

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(10) 国際公開番号

WO 2010/134437 A1

(43) 国際公開日

2010 年 11 月 25 日(25.11.2010)

- (51) 国際特許分類
G06F 12/00 (2006 01)
- (21) 国際出願番号 PCT/JP20 10/057886
- (22) 国際出願日 2010 年 5 月 10 日(10 05 2010)
- (25) 国際出願の言語 日本語
- (26) 国際公開の言語 日本語
- (30) 優先権データ
特願 2009-1 19363 2009 年 5 月 18 日(18 05 2009) JP
特願 2009-218337 2009 年 9 月 23 日(23 09 2009) JP
- (72) 発明者; および
- (71) 出願人 西山 修平(Nishiyama Shuhei) [JP/P], 〒
2790001 千葉県浦安市当麻一丁目2番24号
スプリングガーデン1005 Chiba(JP)
- (81) 指定国(表示のない限り、全ての種類の国内保護が可能) AE, AG, AL, AM, AN, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI,

GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RU, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(84) 指定国(表示のない限り、全ての種類の地域保護が可能) ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), エーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RU, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG)

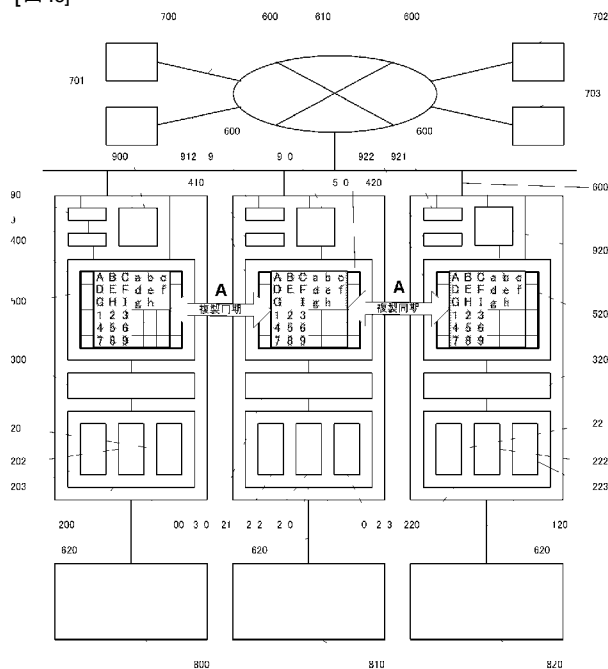
添付公開書類

— 国際索引報告(条約第21条(3))

(54) Title META-INFORMATION SHARING DISTRIBUTED DATABASE SYSTEM IN VIRTUAL SINGLE MEMORY STORAGE

(54) 発明の名称 仮想単一メモリストレージ上におけるメタ情報共有型分散データベース・システム

[図16]



A DISTRIBUTED SHARING

(57) Abstract Concerning scale out of a distributed database system requiring real-time update enabled by means of division among a plurality of database domains of the database system, adaptation to further scale out and higher performance thanks thereto are achieved. A large-scale database system is distributed among a plurality of database domains. In a multi-transaction requiring real-time update of database objects across the database domains, exchange and synchronization of meta-information about the multi transaction, such as status information, are performed, and the partition topology or the replication topology is applied to a database meta-information storage managing unit in a database meta-information management repository device. Thus, the multi-transaction is performed at a high speed.

(57) 要約 【課題】 リアルタイム更新を要件とするデータベース・システムの数多くのデータベース・ドメインへの分割による分散型データベース・システムのスケールアウトに関して、更なる大規模化に対応し、それに伴う更なる高性能化を実現することが本発明の課題である。【解決手段】 本発明は、大規模データベース・システムを前記複数データベース・ドメインに分散し、前記複数データベース・ドメインにまたがるデータベース・オブジェクトのリアルタイム更新をとともうマルチ・トランザクション処理において、そのステータス情報等のメタ情報の交換および同期を、データベース・メタ情報管理リポジトリ装置内のデータベース・メタ

情報格納管理部に対し、パーティショントポロジ技術またはレプリケーショントポロジ技術を適用することによって、更に高速に前記マルチ・トランザクション処理を実施することによって課題を解決する。

明 細 書

発明の名称：

仮想単一メモリストレージ上におけるメタ情報共有型分散データベース・システム

技術分野

[0001] 本発明は、分散データベース・システム及びそれを用いたグリッド・コンピューティング・システムに関する。

背景技術

[0002] 従来の分散データベース技術は、分散トランザクション処理や排他処理にかかわる一貫性および参照整合性制約にかかわる整合性の確保と処理のスループットとのトレードオフの上に成り立っていた。前記一貫性、前記参照整合性の担保のために、従来、二相コミット技術およびレプリケーション技術が適用されてきたが、実用的なスループットを得ることができず、実用に供されて満足な結果を得ているものは寡聞である。一方、二相コミット技術またはレプリケーション技術を使用しない場合、従前技術では、リアルタイム更新の可能な分散データベース・システムは実用化されておらず、リアルタイム更新の可能な分散データベース・システムと称するものは、実際には、単に分散二次記憶媒体上に展開された仮想単一データベース・システムであったりしている。また、インターネット上での検索サイトや仮想店舗サイトでは、更新系はバッチとし分散読み出し特化型分散データベース・システムであったりしている。

[0003] 近年では、データベース管理システムを含むデータベース・ドメインであって、前記複数のデータベース・ドメインをピア・ツー・ピア接続し、データベース・ドメイン内に配置されたデータベースの構造や排他制御、参照性整合等のステータスに関する情報をメタ・データとして本体データとは別に管理し、前記メタ情報を当該メタ情報ごとに前記各データベース・ドメイン間で互いに交換し同期を取ることによりスケーラビリティに優れ、データ

ベース・オブジェクトを任意の粒度で極小化し障害時の障害箇所の局所化と復旧時間の短縮を最小化する分散型データベース・システムに関しては図「に示されるような特許文献「に記載のものが知られている。前記特許文献「では前記メタ情報の交換の具体的な方法までは言及されていないが、前記特許文献「に記載の技術が発明された時点（本件発明者が本件発明の着想を得たのは平成「3年「2月頃であった）では、非特許文献「に記載のパーティショントポロジ技術やレプリケーショントポロジ技術は米国で実用化されたばかり（非特許文献「に記載の当該非特許技術が米国市場に登場し公然実施されたのは平成「4年6月頃であった）で、その時点では米国以外での普及には至っておらず（日本でこの技術の詳細が一部で知られるようになったのは平成「9年7月頃であった）、前記非特許文献「に記載のパーティショントポロジ技術やレプリケーショントポロジ技術を利用して、前記メタ情報の同期を都度個別に交換するのではなく、メタ情報が搭載されたメモリ片ごとに交換して行うという方法は想定されていなかった。前記メタ情報の都度個別交換による同期には、ソフトウェアの構造が複雑になり実装、保守に困難を伴うこととなり、その構造の簡素化が望まれ、また、肥大化するデータが課題とされてきている近年においては、更なる高速化が要望されている。

[0004] また、前記非特許文献「に記載のパーティショントポロジ技術は、図2に示されるように、仮想計算機を含む複数の計算機の主記憶装置上にキャッシュ・トポロジ・システムを構築し、各計算機上の主記憶装置上でメモリ・キャッシュという単位で各々のメモリ・キャッシュが複数の計算機から共有され、比較的大きくない計算機を複数台接続して比較的大規模な仮想計算機を構成するシステムにおいて比較的大規模な仮想インメモリ・データベース・システムを実現する技術として知られている。しかし、これは、分散データベース・システムの実現方法とは言えず、単なる比較的大規模な仮想単一データベース・システムの実現方法であり、障害時にストレージ・レベルでの障害箇所の局所化は出来ても、データベース・システム・レベルとしての障害箇所の局所化とはならず、したがって、前記一貫性、前記参照整合性を

保った状態でのバックアップのためのスナップショットの作成およびロール・フォワードのための更新ジャーナルの収集はデータベース全域に及び、リカバリもデータベース・システム全体に及ぶため、短時間での復旧は容易ではない。

[0005] また、仮想計算機を含む複数の計算機の主記憶装置上にキャッシュ・トポロジ・システムを構築し、各計算機上の主記憶装置上のメモリ・キャッシュをお互いに複製して同期させることによって、他の計算機上の主記憶装置上のデータのアクセスを行う場合でも自身の計算機上の主記憶装置上のデータにアクセスすればよいという技術に関しては非特許文献「の当該ページに記載されているものがレプリケーショントポロジ技術として知られている。しかし、この場合、計算機の台数が増えても、共有するメモリ・キャッシュ量が増加するわけではないので、スケールアウトには寄与しない。

[0006] さらに特許文献4の段落0002には、全ての更新が成功する場合にはコミットを行い、「つでも失敗すればロールバックするような一般的なトランザクションシステムの構成が開示されている。さらに、段落0003には複数のトランザクションを関連させる手法として「入れ子トランザクション」と「連鎖トランザクション」があることが開示されている。段落0004および段落0005には「入れ子トランザクション」に関する詳細な説明が開示されている。また、段落0006および段落0007には「連鎖トランザクション」に関する詳細な説明が開示されている。

先行技術文献

特許文献

[0007] 特許文献1：特許第4「58534号公報

特許文献2：特開第2006-92503号公報

特許文献3：特表第2007-50「456号公報

特許文献4：特許第3732「「3号公報

非特許文献

[0008] 非特許文献1：「分散インメモリキャッシュとデータグリッド」，[季刊]ユ

ニックス・マガジン, (株)アスキー・メディアワークス, 2009年4月号,
p. 72-77

非特許文献2: 「スケールアウトの技術」, [季刊] ユニックス・マガジン,
(株)アスキー・メディアワークス, 2009年4月号, p. 78-9「

発明の概要

発明が解決しようとする課題

[0009] 複数のデータベース管理装置を複数のデータベース・ドメインに分け、各々のデータベース・ドメイン相互がそのデータベース・ドメイン内のデータベース管理装置内の各データベース・オブジェクトのステータス等のメタ情報を交換する仕組みを各々のデータベース・ドメイン毎に持つことにより、前記二相コミット技術および前記レプリケーション技術の適用なしに、実用スループットを担保するリアルタイム更新処理が可能な分散型データベース・システムを実現する方法が特許文献1に記載されている。しかし、前記メタ情報の都度個別交換による同期には、ソフトウェアの構造が複雑になり実装、保守に困難を伴うこととなり、その構造の簡素化が望まれ、また、肥大化するデータが課題とされてきている近年においては、更なる高速化が必要とされている。

[0010] 単一表（以下リレーション）が所有する行（以下タプル）が大量となった際、スケールアウトするための方法として、タプルの中の一部データを引数とするコンシステント・ハッシングによってタプル群を水平分割する技術が非特許文献2に記載されている。そこでは、データベース・オブジェクトの物理的な配置先である物理ノードの追加、離脱に関する柔軟な対応方法が今後の課題であるとされている。一方、前記スケールアウトに対し、ハッシングによってタプル群を水平分割するにあたって、分割されたタプル群に対して直接物理ノードを対応させるのではなく、論理ノードとしてのデータベース・オブジェクト・セットを対応させ、データベース・オブジェクト・セットと物理ノード、すなわち論理ノードと物理ノードをマッピングさせるハッシングとマッピングのハイブリッド方式による前記物理ノードの追加、離脱に

関する柔軟性の確保に関する技術が、非特許文献 2 に記載のコンシステント・ハッシングを利用する技術に先行して特許文献 2 に記載されている。図 20 は特許文献 2 に記載の論理ノードと物理ノードをマッピングさせるハッシングとマッピングのハイブリット方式を実装するデータベース・オブジェクト管理装置特定装置 907 の詳細構成図である。

特許文献 2 に記載の大量タプルを持つリレーシヨンの水平分割に関しては、単一の前記データベース・ドメイン内での分散データベース・システムを想定しているため、更なるスケールアウトのために、複数データベース・ドメインにまたがる前記リレーシヨンの水平分割にも対応することが必要とされている。

図 2「は複数データベース・ドメインにまたがる前記リレーシヨンの水平分割にも対応するための請求項 4 に記載のデータベース・オブジェクト管理装置特定装置 907 とデータベース・メタ情報管理リポジトリ装置 500 の詳細構成図である。

[0011] 単一の前記データベース・ドメイン内でのマルチ・トランザクション処理の一貫性を一相コミットで確保する方法が特許文献 3 に記載されている。特許文献 3 によれば、複数の前記データベース・ドメイン相互でのマルチ・トランザクション処理の一貫性の保持のためには二相コミットで確保しなければならない旨が記載されており、複数データベース・ドメインにまたがる一貫性が保持されたマルチ・トランザクション処理の実用的なスループットの確保は難しい。また、マルチ・トランザクション処理一貫性を確保する方法として、前記マルチ・トランザクション処理を構成する各単一トランザクション処理のステータスは、該ステータスのステータス情報を該単一トランザクション毎に保持することになっており、非常に複雑なものになっている。

[0012] 非特許文献「に記載のパーティショントポロジ技術のデータベース管理システムの大規模化およびインメモリ化による高速化に寄与し、当該データベース管理システムは、複数の比較的小型な計算機を複数台接続して大規模

単一インスタンス仮想データベースを構成するため、前記一貫性および前記整合性に対する前記二相コミット技術およびレプリケーション技術の適用を必要としない。しかし、該前記大規模単一インスタンス仮想データベースを構成する複数台の前記比較的小型な計算機上のデータの集合の各々と該前記大規模単一インスタンス仮想データベースを構成するリレーションやタプル毎のようなデータベース・オブジェクトの意味のある集合とは、直接には対応しない。したがって、特定の前記データベース・オブジェクトを特定のノードである前記比較的小型な計算機に配置し管理することができない。これは、該特定のデータベース・オブジェクトを頻繁にアクセスする前記ノードである比較的小型な計算機に該特定のデータベース・オブジェクトを配置し、別の特定のデータベース・オブジェクトを頻繁にアクセスする前記ノードである比較的小型な計算機に該別の特定のデータベース・オブジェクトを配置し、前記ノードである比較的小型な計算機と前記特定のデータベース・オブジェクトとを明示的に関連付けて管理することができないということを意味している。一方、非特許文献2に記載されているように、特定のノードである比較的管理しやすい比較的小型な計算機上に管理しやすい大きさに分割された前記データベース・オブジェクト群を配置し、分割された前記データベース・オブジェクト相互の緩やかに疎結合される形の大規模分散データベース・システムというものが、分割された前記データベース・オブジェクト群の独立性のために必要とされている。これは、分割された前記データベース・オブジェクト群の保守性、秘匿性に大きく寄与する。しかし、前記分割された前記データベース・オブジェクト間の一貫性および整合性の保持に関しては前記二相コミット技術およびレプリケーション技術によらなければならない、性能面に問題を残していた。さらに、非特許文献2に記載されているように、大量のタプルを所有する「個のリレーションにおいて、各々のタプルにユニークなIDを付加して他のタプルと識別するとき、そのタプルの該IDと物理格納装置を持つ計算機との関係付けを、前記物理格納装置を持つ計算機のシステムへの参加、離脱（故障等による強制的なものも含む）に対

しても柔軟にダイナミックに対応されていることが必要とされている。

- [0013] したがって、リアルタイム更新を要件とするデータベース・システムの複数のデータベース・ドメインへの分割による分散型データベース・システムのスケールアウトに関して、更なる複雑化に対応し、更なる高性能化を実現することが本発明の課題である。

課題を解決するための手段

- [0014] 本発明は、大規模データベース・システムを前記複数データベース・ドメインに分散し、前記複数データベース・ドメインにまたがるデータベース・オブジェクトのリアルタイム更新をとまなうマルチ・トランザクション処理において、データ・ディクショナリ情報、排他制御情報、参照整合性情報等の各種ステータス情報および更新アクセス・カウンタおよびデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブル等のメタ情報の交換および同期を、非特許文献「に記載のパーティショントポロジ技術またはレプリケーショントポロジ技術を適用することによって課題を解決する。

- [0015] また、本発明では一貫性が保持されなければならない一連のマルチ・トランザクションを識別するためのIDとして更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別されることを特徴とする更新アクセス・カウンタを使用している。すなわち、本発明は、特許文献3に記載のコーディネータによる各トランザクションのステータスの一元共有による一段階コミット・プロトコルによるマルチ・トランザクション処理に代え、マルチ・トランザクション処理の起点となるプロセスのプロセスIDによって識別されるカウンタのアップダウンにより、カウンタが初期値に戻ったときをひとつの一貫性が保持されなければならない一連のマルチ・トランザクション処理の完了とみなすことにするという更新アクセス・カウンタ技術を適用することによって課題を解決する。

発明の効果

- [0016] 特許文献「に記載の分散型データベース・システムにおける当該メタ情

報の共有の実装・保守を更に容易にし、また、当該メタ情報を持つ複数のメモリ・キャッシュの高速な同期が可能なため、個別にメタ情報の同期を取る必要があった特許文献「に記載の分散データベース・システムを、性能面においても勝ることになる。

また、複数のデータベースにまたがるマルチ・トランザクション処理を、オーバーヘッドの大きな二相コミット・プロトコルによらない方法で、シンプルに高速に実施することが可能とし、さらに、複数のデータベースが疎結合することによって有機的に結びついて大規模なデータベース・システムを構成するようなシステムにおいても、前記トランザクション起動プロセス単位のアクセス・カウンタの共有のみで、簡単に分散マルチ・トランザクション処理が可能となる。

[0017] 特許文献2に記載のハッシング・マッピング・ハイブリッド方式による大量タプルを有するリレーションの水平分散を効率的に、前記データベース・ドメインにまたがって実施できることになる。

[0018] 特許文献3に記載のコーディネータによる各トランザクションのステータスの一元共有による前記ドメイン内における一段階コミットプロトコルによるマルチ・トランザクション処理に代え、複数のデータベース・ドメインにまたがるマルチ・トランザクション処理を、その起点となるプロセスのプロセスIDによって認識されるカウンタのアップダウンにより、カウンタが初期値に戻ったときをひとつのマルチ・トランザクション処理の完了とみなすことにより、複数のデータベース・ドメインにまたがるマルチ・トランザクション処理をシンプルに高速に実施できることになる。

図面の簡単な説明

[0019] [図1] 図「は特許文献「に記載の分散型データベース・システムの構成を示した構成図である。

[図2] 図2は非特許文献「に記載の技術により構成することができる仮想単一主記憶装置ストレージ上の仮想単一データベース・システムの構成を示した構成図である。

[図3] 図3はパーティショントポロジ技術を適用した請求項1に記載の分散型データベース・システムの構成を示した構成図である。(実施例1)

[図4] 図4はレプリケーショントポロジ技術を適用した請求項1に記載の分散型データベース・システムの構成を示した構成図である。(実施例2)

[図5] 図5はパーティショントポロジ技術を適用した請求項3に記載の分散型データベース・システムの構成を示した構成図である。(実施例3)

[図6] 図6はレプリケーショントポロジ技術を適用した請求項3に記載の分散型データベース・システムの構成を示した構成図である。(実施例4)

[図7] 図7はパーティショントポロジ技術を適用した請求項4に記載の分散型データベース・システムの構成を示した構成図である。(実施例5)

[図8] 図8はレプリケーショントポロジ技術を適用した請求項4に記載の分散型データベース・システムの構成を示した構成図である。(実施例6)

[図9] 図9はパーティショントポロジ技術を適用した請求項5に記載の分散型データベース・システムの構成を示した構成図である。(実施例7)

[図10] 図10はレプリケーショントポロジ技術を適用した請求項5に記載の分散型データベース・システムの構成を示した構成図である。(実施例8)

[図11] 図11はパーティショントポロジ技術を適用した請求項6に記載の分散型データベース・システムの構成を示した構成図である。(実施例9)

[図12] 図12はレプリケーショントポロジ技術を適用した請求項6に記載の分散型データベース・システムの構成を示した構成図である。(実施例10)

[図13] 図13はパーティショントポロジ技術を適用した請求項7に記載の分散型データベース・システムの構成を示した構成図である。(実施例11)

[図14] 図14はレプリケーショントポロジ技術を適用した請求項7に記載の分散型データベース・システムの構成を示した構成図である。(実施例12)

[図15] 図15はパーティショントポロジ技術を適用した請求項8に記載の分散型データベース・システムの構成を示した構成図である。(実施例13)

[図16] 図 1 6 はレプリケーショントポロジ技術を適用した請求項 8 に記載の分散型データベース・システムの構成を示した構成図である。（実施例 1 4）

[図17] 図 1 7 は請求項 2 に記載の共有カウンタ方式によるマルチ・トランザクション処理において全ての更新処理が成功した場合を例示したシーケンス図である。

[図18] 図 1 8 は請求項 2 に記載の共有カウンタ方式によるマルチ・トランザクション処理において途中で更新処理が失敗した場合を例示したシーケンス図である。

[図19] 図 1 9 は請求項 2 に記載の分散マルチ・トランザクション処理装置とデータベース・メタ情報管理リポジトリ装置の詳細構成図である。

[図20] 図 2 0 は特許文献 2 に記載のデータベース・オブジェクト管理装置特定装置の詳細構成図である。

[図21] 図 2 1 は請求項 5 に記載のデータベース・オブジェクト管理装置特定装置とデータベース・メタ情報管理リポジトリ装置の詳細構成図である。

[図22] 図 2 2 は請求項 2 に記載の共有カウンタ方式による分散マルチ・トランザクション処理装置の一例である。

発明を実施するための形態

[0020] 以下、本発明の実施の形態について、図を用いて説明する。なお、本発明は、これら実施の形態に何ら限定されるものではなく、その要旨を逸脱しない範囲において、種々なる態様で実施し得る。

特に本発明に関わるデータベース管理システムについて、リレーショナル・データベース、オブジェクト・データベースとその型に依存するものではないが、説明を容易にするために、以下の実施例ではリレーショナル・データベース管理システムによって例示する。ここで、リレーショナル・データベースの定義は、「タプルとアトリビュートの外積をリレーションとする」に拠る。

実施例 1

[0021] 図3を使って、パーティショントポロジ技術を適用した請求項「に記載の分散型データベース・システムについて説明する。

ネットワーク600^Aに配置された複数のデータベース・ドメイン「00、
「0、「20の各々の前記データベース・ドメイン内にあって、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置300、3「0、320と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置200、2「0、220と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置400、4「0、420と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

自身のデータベース・ドメイン内のデータベース・オブジェクト管理装置が保持する管理下のデータベースの構造を定義するデータ・ディクショナリ情報と、

自身のデータベース・ドメイン内のデータベース・オブジェクト管理装置内のデータベース・オブジェクトが他の前記データベース・ドメイン内のデ

データベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対して外部参照状態にある時の参照被参照状態にある双方のデータベース・オブジェクトに関するステータス情報と、

自身の前記データベース・ドメイン内の前記データベース・オブジェクト管理装置が管理する前記データベース・オブジェクトを他のデータベース・ドメインを管理する前記データベース・メタ情報管理リポジトリ装置を直接または間接的にアクセスする端末計算機またはサーバ計算機からの読み出し状態にあり共有ロック状態にある当該データベース・オブジェクトのステータス情報と、

自身の前記データベース・ドメイン内の前記データベース・オブジェクト管理装置が管理する前記データベース・オブジェクトを他のデータベース・ドメインを管理する前記データベース・メタ情報管理リポジトリ装置を直接または間接的にアクセスする端末計算機またはサーバ計算機からの書き込み状態にあり排他ロック状態にある当該データベース・オブジェクトのステータス情報と、

を分散型データベース・システムのデータベース・メタ情報として格納し、

自身のデータベース・ドメイン内で処理が進行し、前記データベース・オブジェクトに関するステータス情報に変化が及ぼす事象が発生した場合には速やかに当該データベース・オブジェクトに関するステータス情報を更新管理する、

データベース・メタ情報格納管理部 500、510、520と、

を有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記デー

データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 2

[0022] 図4を使って、レプリケーショントポロジ技術を適用した請求項「に記載の分散型データベース・システムについて説明する。

ネットワーク600^Aに配置された複数のデータベース・ドメイン「00、「10、「20の各々の前記データベース・ドメイン内にあって、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置300、310、320と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置200、210、220と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置400、410、420と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するための

データベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

自身のデータベース・ドメイン内のデータベース・オブジェクト管理装置が保持する管理下のデータベースの構造を定義するデータ・ディクショナリ情報と、

自身のデータベース・ドメイン内のデータベース・オブジェクト管理装置内のデータベース・オブジェクトが他の前記データベース・ドメイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対して外部参照状態にある時の参照被参照状態にある双方のデータベース・オブジェクトに関するステータス情報と、

自身の前記データベース・ドメイン内の前記データベース・オブジェクト管理装置が管理する前記データベース・オブジェクトを他のデータベース・ドメインを管理する前記データベース・メタ情報管理リポジトリ装置を直接または間接的にアクセスする端末計算機またはサーバ計算機からの読み出し状態にあり共有ロック状態にある当該データベース・オブジェクトのステータス情報と、

自身の前記データベース・ドメイン内の前記データベース・オブジェクト管理装置が管理する前記データベース・オブジェクトを他のデータベース・ドメインを管理する前記データベース・メタ情報管理リポジトリ装置を直接または間接的にアクセスする端末計算機またはサーバ計算機からの書き込み状態にあり排他ロック状態にある当該データベース・オブジェクトのステータス情報と、

を分散型データベース・システムのデータベース・メタ情報として格納し、

自身のデータベース・ドメイン内で処理が進行し、前記データベース・オブジェクトに関するステータス情報に変化が及ぼす事象が発生した場合には速やかに当該データベース・オブジェクトに関するステータス情報を更新管理する、

データベース・メタ情報格納管理部 500、510、520と、

を有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記 分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されることを特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 3

[0023] 図5および図「9」を使って、パーティショントポロジ技術を適用した請求項3に記載の分散型データベース・システムについて説明する。

ネットワーク6007に配置された複数のデータベース・ドメイン「00」、「10」、「20」の各々の前記データベース・ドメイン内にあって、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置300、310、320と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置200、210、220と、

分散マルチ・トランザクション処理においてトランザクション処理一貫性を保持しながら実行するための分散マルチ・トランザクション処理装置900

、 910、920と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置400、410、420と、

で構成されるメタ情報共有分散型データベース・システムであって、

分散マルチ・トランザクション処理装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求される、自身または他の前記データベース・ドメイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトランザクション処理において、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ903が存在するものとし、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、

該起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドが存在すればそのプロセスないしスレッドによって要求される更新アクセスが発生する都度、該前記更新アクセスを要求する起点となったプロセスを識別するプロ

セス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを一定数ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致した時に、該前記起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのコミット処理を実行し、また、

実際の更新アクセスが失敗したという情報を一個でも受け取った場合、余の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点となったプロセス自身および該前記起点となったプロセスないしスレッドが存在するならばそのプロセスないしスレッドによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのロールバック処理を実行する、

ことを特徴とする分散マルチ・トランザクション処理実行部 906 と、
を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ

を有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機で

ある各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置中にあるデータベース・メタ情報格納管理部510、520とが、前記分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

図「7は共有カウンタ方式によるマルチ・トランザクション処理において全ての更新処理が成功した場合を例示したシーケンス図である。

図「8は共有カウンタ方式によるマルチ・トランザクション処理において途中で更新処理が失敗した場合を例示したシーケンス図である。

実施例 4

[0024] 図6および図「9を使って、レプリケーショントポロジ技術を適用した請求項3に記載の分散型データベース・システムについて説明する。

ネットワーク6007に配置された複数のデータベース・ドメイン「00、「10、「20の各々の前記データベース・ドメイン内にあって、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置300、310、320と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置200、210、220と、

分散マルチ・トランザクション処理においてトランザクション処理一貫性を保持しながら実行するための分散マルチ・トランザクション処理装置900、910、920と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置400、410、420と、

で構成されるメタ情報共有分散型データベース・システムであって、

分散マルチ・トランザクション処理装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求される、自身または他の前記データベース・ドメイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトランザクション処理において、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ903が存在するものとし、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、

該起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドが存在すればそのプロセスないしスレッドによって要求される更新アクセスが発生する都度、該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウント

を一定数ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致した時に、該前記起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのコミット処理を実行し、また、

実際の更新アクセスが失敗したという情報を一個でも受け取った場合、余の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点となったプロセス自身および該前記起点となったプロセスないしスレッドが存在するならばそのプロセスないしスレッドによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのロールバック処理を実行する、

ことを特徴とする分散マルチ・トランザクション処理実行部 906 と、
を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ

を有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、前記メモ

リ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

図「7は共有カウンタ方式によるマルチ・トランザクション処理において全ての更新処理が成功した場合を例示したシーケンス図である。

図「8は共有カウンタ方式によるマルチ・トランザクション処理において途中で更新処理が失敗した場合を例示したシーケンス図である。

実施例 5

[0025] 図7および図「9を使って、パーティショントポロジ技術を適用した請求項4に記載の分散型データベース・システムについて説明する。

実施例「に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

分散マルチ・トランザクション処理においてトランザクション処理一貫性を保持しながら実行するための分散マルチ・トランザクション処理装置900、910、920と、

を構成要素とするメタ情報共有分散型データベース・システムであって、

分散マルチ・トランザクション処理装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが「00、「10、「20のように複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクト

を管理する前記データベース・オブジェクト管理装置 300、310、320を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求される、自身または他の前記データベース・ドメイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトランザクション処理において、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ 903が存在するものとし、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、該起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドが存在すればそのプロセスないしスレッドによって要求される更新アクセスが発生する都度、該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを一定数ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致した時に、該前記起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッド

によって実行されたすべての更新アクセスのコミット処理を実行し、また、
実際の更新アクセスが失敗したという情報を一個でも受け取った場合、余
の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点とな
ったプロセス自身および該前記起点となったプロセスないしスレッドが存在
するならばそのプロセスないしスレッドによって直接または間接的に引き起
こされた自身または他のデータベース・ドメイン内の端末計算機またはサー
バ計算機上のプロセスないしスレッドによって実行されたすべての更新アク
セスのロールバック処理を実行する、
ことを特徴とする分散マルチ・トランザクション処理実行部 906 と、
を有し、

データベース・メタ情報管理リポジトリ装置 400、410、420 は、
ここではネットワーク上には互いに通信が可能なデータベース・ドメイン
が複数個あるものとして、
複数のデータベース・ドメインから要求される「または2以上の前記更新ア
クセスの起点となるプロセスを識別するプロセス識別子によって識別される
更新アクセス・カウンタ

を有すること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

図 17 は共有カウンタ方式によるマルチ・トランザクション処理において全
ての更新処理が成功した場合を例示したシーケンス図である。

図 18 は共有カウンタ方式によるマルチ・トランザクション処理において途
中で更新処理が失敗した場合を例示したシーケンス図である。

実施例 6

[0026] 図 8 および図 9 を使って、レプリケーショントポロジ技術を適用した
請求項 4 に記載の分散型データベース・システムについて説明する。

実施例 2 に記載のメタ情報共有型分散データベース・システムにおいて、
さらに、
分散マルチ・トランザクション処理においてトランザクション処理一貫性を

保持しながら実行するための分散マルチ・トランザクション処理装置 900、910、920と、

を構成要素とするメタ情報共有分散型データベース・システムであって、

分散マルチ・トランザクション処理装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが「00、「10、「20のように複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置300、310、320を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求される、自身または他の前記データベース・ドメイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトランザクション処理において、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ903が存在するものとし、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、

該起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドが存在すればそのプロセスないしスレッドによって要求される更新アクセスが発生する都度、該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを一定数ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致した時に、該前記起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのコミット処理を実行し、また、

実際の更新アクセスが失敗したという情報を一個でも受け取った場合、余の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点となったプロセス自身および該前記起点となったプロセスないしスレッドが存在するならばそのプロセスないしスレッドによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのロールバック処理を実行する、

ことを特徴とする分散マルチ・トランザクション処理実行部 906 と、

を有し、

データベース・メタ情報管理リポジトリ装置 400、410、420 は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ

を有すること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

図「7 は共有カウンタ方式によるマルチ・トランザクション処理において全ての更新処理が成功した場合を例示したシーケンス図である。

図「8 は共有カウンタ方式によるマルチ・トランザクション処理において途

中で更新処理が失敗した場合を例示したシーケンス図である。

実施例 7

[0027] 図9および図2「を使って、パーティショントポロジ技術を適用した請求項5に記載の分散型データベース・システムについて説明する。

ネットワーク600Aに配置された複数のデータベース・ドメイン「00、
「10、「20の各々の前記データベース・ドメイン内にあって、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置300、310、320と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置200、210、220と、

単一の前記データベース・オブジェクト格納装置中に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置中に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置907と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置400、410、420と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、
ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース

・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 901、911、921と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、912、922と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する」「または2以上の前記データベース・オブジェクト格納装置上の」「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理 リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブル 9 05

を有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理 リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理 リポジトリ装置の前記データベース・メタ情報格納管理部 5 0 0と、他のデータベース・ドメイン内の前記データベース・メタ情報管理 リポジトリ装置内にあるデータベース・メタ情報格納管理部 5 「 0、5 2 0とが、前記 分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理 リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 8

[0028] 図 1 および図 2 を使って、レプリケーショントポロジ技術を適用した請求項 5 に記載の分散型データベース・システムについて説明する。

ネットワーク 600 に配置された複数のデータベース・ドメイン 100、110、120 の各々の前記データベース・ドメイン内にあって、

データベースを管理する「または 2 以上のデータベース・オブジェクト管理装置 300、310、320 と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または 2 以上のデータベース・オブジェクト格納装置 200、210、220 と、

単一の前記データベース・オブジェクト格納装置中に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または 2 以上の前記データベース・オブジェクト管理装置が管理する「または 2 以上の前記データベース・オブジェクト格納装置中に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置 907 と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または 2 以上のデータベース・メタ情報管理リポジトリ装置 400、410、420 と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクト

トの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 901、911、921と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、912、922と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する」または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブル 905

を有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記 分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 9

[0029] 図 1 および図 2 を使って、パーティショントポロジ技術を適用した請求項 6 に記載の分散型データベース・システムについて説明する。

実施例 1 に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置 200、210、220 内に収容しきれないタブル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の 1 または 2 以上の前記データベース・オブジェクト管理装置 300、310、320 が管理する 1 または 2 以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置 907 と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための 1 または 2 以上のデータベース・メタ情報管理リポジトリ装置 400、410、420 と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、

ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 901、911、921 と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベー

ス・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、912、922と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブ

ジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記 分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 10

[0030] 図「2 および図2」を使って、レプリケーショントポロジ技術を適用した請求項6に記載の分散型データベース・システムについて説明する。

実施例2に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置 2 0 0、2 1 0、2 2 0 内に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または 2 以上の前記データベース・オブジェクト管理装置 3 0 0、3 1 0、3 2 0 が管理する「または 2 以上の前記データベース・オブジェクト格納装置中に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置 9 0 7 と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または 2 以上のデータベース・メタ情報管理リポジトリ装置 4 0 0、4 1 0、4 2 0 と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、
ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 9 0 1、9 1 1、9 2 1 と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 9 0 2、9 1 2、9 2 2 と、

を有し、

データベース・オブジェクト検索加工装置は、
該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、
前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、
を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該

前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、または、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記 分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 11

[0031] 図13および図21を使って、パーティショントポロジ技術を適用した請求項7に記載の分散型データベース・システムについて説明する。

実施例3に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置200、210、220内に収容しきれないタブル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置300、310、320が管理する「または2以上の前

記データベース・オブジェクト格納装置内に收容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置 907 と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置 400、410、420と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、
ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 901、911、921と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、912、922と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する「ま

たは2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該デ

データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 12

[0032] 図「4 および図2」を使って、レプリケーショントポロジ技術を適用した請求項7に記載の分散型データベース・システムについて説明する。

実施例4に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置200、210、220内に収容しきれないタブル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置300、310、320が管理する」または2以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置907と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理

リポジトリ装置 400、410、420と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、

ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 901、911、921と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、912、922と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する」または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すな

わちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン

内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部5「0、520とが、前記分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

実施例 13

[0033] 図「5および図2」を使って、パーティショントポロジ技術を適用した請求項7に記載の分散型データベース・システムについて説明する。

実施例5に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置200、2「0、220内に収容しきれないタブル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置300、3「0、320が管理する「または2以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置907と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置400、4「0、420と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、

ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定する

ためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 901、911、921と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、912、922と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する」または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータ・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメイン

が複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記 分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

[0034] 図「5は、本発明の「実施例の構成図であって、データベース・ドメイン「00を一台のサーバ計算機によって実現しようとしたものであって、データベース・ドメイン「00内には、データベース・オブジェクト管理装置

特定装置 907 (図「5 中省略」) 内に配置されたデータベース・オブジェクト・セット識別記号変換部 901 およびデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、分散マルチ・トランザクション処理装置 906 (図「5 中省略」) 内に配置された分散マルチ・トランザクション処理実行部 900、データベース・メタ情報管理リポジトリ装置 400 内に配置されたデータベース・メタ情報格納管理部 500、データベース・オブジェクト管理装置 300、データベース・オブジェクト格納装置 200 が配置されており、障害時の復旧のためのスナップショットおよびジャーナルを格納するストレージ 800 を持つストレージ・サーバ計算機とネットワーク 600 で接続されており、データベース・ドメイン「00」と同様の構造のデータベース・ドメイン「10」およびデータベース・ドメイン「20」と「AN600」で接続し、非特許文献「」に記載のパーティショントポロジ技術によりデータベース・メタ情報格納管理部 500 内のパーティション・キャッシュを分散共有し、メタ情報の同期により、高速にリアルタイム更新可能な分散データベースを稼動する。

[0035] ここで、図「5」に記載のストレージ 800 はデータベース・ドメイン「00」と WAN を介して物理的に遠距離に存在するものであってもよい。

[0036] ここで、図「5」に記載のストレージ 800 はデータベース・ドメイン「00」を構成する一台のサーバ計算機に制御される ATA、FC 等のインターフェース・ケーブルで接続された二次記憶装置でもよい。

[0037] ここで、図「5」に記載のリレーション 201、202、203、204、205、206、207、208、209 は、単一のリレーションに関する大量のタプル群を水平分割したタプル群のいずれかを含んでもよく、該タプル群は、リレーション 201 およびリレーション 203 のような同一データベース・ドメインに含まれるリレーションに含まれてもよく、また、リレーション 201 およびリレーション 205 のような異なるデータベース・ドメインに配置されたリレーションに含まれてもよい。

[0038] 単一のリレーションに関する大量のタプル群が特定のアルゴリズムで水

平分割されたタプル群はデータベース・オブジェクト・セットと定義され、データベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを使用して、物理的な格納先のロケーションが与えられるが、格納先の格納領域の利用状況のバランスを調整するために、動的に再配置できるようになっていてもよい。

実施例 14

[0039] 図「6 および図2」を使って、レプリケーショントポロジ技術を適用した請求項8に記載の分散型データベース・システムについて説明する。

実施例6に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置200、210、220内に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置300、310、320が管理する「または2以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置907と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置400、410、420と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、

ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定する

ためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部 901、911、921と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部 902、912、922と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する」または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータ・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメイン

が複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部500と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部510、520とが、前記 分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システムであってもよい。

[0040] 図「6は、本発明の「実施例の構成図であって、データベース・ドメイ

ン「00を一台のサーバ計算機によって実現しようとしたものであって、データベース・ドメイン「00内には、データベース・オブジェクト管理装置特定装置907（図「5中省略）内に配置されたデータベース・オブジェクト・セット識別記号変換部901およびデータベース・オブジェクト・セット識別記号ロケーション情報変換部902、分散マルチ・トランザクション処理装置906（図「5中省略）内に配置された分散マルチ・トランザクション処理実行部900、データベース・メタ情報管理リポジトリ装置400内に配置されたデータベース・メタ情報格納管理部500、データベース・オブジェクト管理装置300、データベース・オブジェクト格納装置200が配置されており、障害時の復旧のためのスナップショットおよびジャーナルを格納するストレージ800を持つストレージ・サーバ計算機とネットワーク600で接続されており、データベース・ドメイン「00と同様の構造のデータベース・ドメイン「10およびデータベース・ドメイン「20と」AN600で接続し、非特許文献「に記載のレプリケーションポロジ技術によりデータベース・メタ情報格納管理部500内のレプリケーション・キャッシュを複製共有し、メタ情報の同期により、高速にリアルタイム更新可能な分散データベースを稼動する。

- [0041] ここで、図「6に記載のストレージ800はデータベース・ドメイン「00とWANを介して物理的に遠距離に存在するものであってもよい。
- [0042] ここで、図「6に記載のストレージ800はデータベース・ドメイン「00を構成する一台のサーバ計算機に制御されるATA、FC等のインターフェース・ケーブルで接続された二次記憶装置でもよい。
- [0043] ここで、図「6に記載のリレーション201、202、203、204、205、206、207、208、209は、単一のリレーションに関する大量のタプル群を水平分割したタプル群のいずれかを含んでもよく、該タプル群は、リレーション201およびリレーション203のような同一データベース・ドメインに含まれるリレーションに含まれてもよく、また、リレーション201およびリレーション205のような異なるデータベース・ド

メインに配置されたリレーションに含まれてもよい。

- [0044] 単一のリレーションに関する大量のタプル群が特定のアルゴリズムで水平分割されたタプル群はデータベース・オブジェクト・セットと定義され、データベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを使用して、物理的な格納先のロケーションが与えられるが、格納先の格納領域の利用状況のバランスを調整するために、動的に再配置できるようになっていてもよい。

実施例 15

- [0045] 図9および図2「を使って、パーティショントポロジ技術を適用した請求項9に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする実施例7に記載のメタ情報共有型分散データベース・システムであってもよい。

さらに、図10および図2「を使って、レプリケーショントポロジ技術を適用した請求項8に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする実施例9に記載のメタ情報共有型分散データベース・システムであってもよい。

実施例 16

[0046] 図「および図2」を使って、パーティショントポロジ技術を適用した請求項「0に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することの特徴とする実施例9に記載のメタ情報共有型分散データベース・システムであってもよい。

さらに、図「2および図2」を使って、レプリケーショントポロジ技術を適用した請求項「0に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することの特徴とする実施例「0に記載のメタ情報共有型分散データベース・システムであってもよい。

実施例 17

[0047] 図「3および図2」を使って、パーティショントポロジ技術を適用した請求項「1に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することの特徴とする実施例「1に記載のメタ情報共有型分散データベース・システムであ

ってもよい。

図「4 および図2」を使って、レプリケーショントポロジ技術を適用した請求項「1」に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする実施例「2」に記載のメタ情報共有型分散データベース・システムであってもよい。

実施例 18

[0048] 図「5 および図2」を使って、パーティショントポロジ技術を適用した請求項「2」に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする実施例「3」に記載のメタ情報共有型分散データベース・システムであってもよい。

図「6 および図2」を使って、レプリケーショントポロジ技術を適用した請求項「2」に記載の分散型データベース・システムについて説明する。

データベース・オブジェクト管理装置特定装置907は、データベース・オブジェクト・セット識別記号ロケーション情報変換部902、912、922のロケーション情報を変更する機能を有し、データベース・オブジェクト格納装置200、210、220の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする実施例「4」に記載のメタ情報共有型分散データベース・システムであ

ってもよい。

実施例 19

[0049] 図「7 および図「8 および図「9 および図22を使って、アクセス・カウンタ技術を適用した請求項2に記載の分散マルチ・トランザクション制御システムについて説明する。

ネットワーク900^Aに配置された、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置300と、

前記データベース・オブジェクト管理装置300が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置200と、分散マルチ・トランザクション処理においてトランザクション処理一貫性を保持しながら実行するための分散マルチ・トランザクション処理装置900と、

前記データベースの持つメタ情報を管理するための「または2以上のデータベース・メタ情報管理リポジトリ装置400と、で構成されるアクセス・カウンタ式分散マルチ・トランザクション制御システムであって、

データベース・メタ情報管理リポジトリ装置400は、

「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別されることを特徴とする更新アクセス・カウンタ903と、

を有し、

図「9に記載されているように、

分散マルチ・トランザクション処理装置900は、

直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求されるデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトラ

ンザクション処理において、

図「7」に記載されているように、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、
該起点となったプロセス自身および該前記起点となったプロセスによって
直接または間接的に引き起こされた自身のデータベース・ドメイン内の端末
計算機またはサーバ計算機上のプロセスないしスレッドが存在すればそのプ
ロセスないしスレッドによって要求される更新アクセスが発生する都度、該
前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別
子によって識別される該前記更新アクセス・カウンタ内のカウントを一定数
ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新
アクセスを要求する起点となったプロセスを識別するプロセス識別子によっ
て識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と
同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致し
た時に、該前記起点となったプロセス自身および該前記起点となったプロセ
スによって直接または間接的に引き起こされた自身または他のデータベース
・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッド
によって実行されたすべての更新アクセスのコミット処理を実行し、

また、図「8」に記載されているように、

実際の更新アクセスが失敗したという情報を一個でも受け取った場合、余
の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点とな
ったプロセス自身および該前記起点となったプロセスないしスレッドが存在
するならばそのプロセスないしスレッドによって直接または間接的に引き起
こされた自身または他のデータベース・ドメイン内の端末計算機またはサー
バ計算機上のプロセスないしスレッドによって実行されたすべての更新アク
セスのロールバック処理を実行する、

分散マルチ・トランザクション処理実行部 906 と、

を有する、

ことを特徴とする、アクセス・カウンタ式分散マルチ・トランザクション制御システムであつてもよい。

産業上の利用可能性

[0050] 大規模データベース・システムを、独立したデータベースとして機能する極限にまで細分化し、それらの細分化された独立データベースが疎結合で再統合することにより、同規模の大規模分散データベース・システムが提供される。極小化された独立データベースは、近年の主記憶装置の大容量化、低価格化と相俟って、シリアライズ、デシリアライズといった時間のかかる重い処理をまったく含まないインメモリ・データベースとして低価格なサーバ装置上に搭載される。この安価なデータベース・サーバ装置をインフィニ・バント等の高速ネットワーク技術により多数接続し、データベース・サーバ装置をダイナミックに追加したり、また、一部のデータベース・サーバ装置に障害が発生し切り離さざるを得なくなつたときにも、その影響を局所化し、障害が取り除かれた時には、局所化故に、ロール・フォワードの起点となるスナップショットおよびジャーナルも極小化され、短時間でリカバリできるリアルタイム更新が可能なグリッド・データベース・システムを提供することができる。

前記データベース・ドメインの構成を「個の単位とするプラグイン・データベース・サーバ・アプライアンスを構成し、データセンタのラック内に前記プラグイン・データベース・サーバ・アプライアンスを配置し、接続し、スケールアウト可能な大規模グリッド・データベース・システムを提供することができる。

符号の説明

[0051] 「00、」「0、」「20 データベース・ドメイン
 200、2「0、220 データベース・オブジェクト格納装置
 201、202、203、211、212、213、221、222、2
 23 リレーション
 300、3「0、320 データベース・オブジェクト管理装置

- 4 0 0、4 「 0、4 2 0 データベース・メタ情報管理 リポジトリ装置
- 5 0 0、5 「 0、5 2 0 データベース・メタ情報格納管理部
- 6 0 0 LAN、WANのTCP/IPネットワーク
- 6 「 0 インターネット網 7 0 0、7 「 0、7 2 0、7 3 0 端末装置
- 8 0 0、8 「 0、8 2 0 ストレージ
- 9 0 0、9 「 0、9 2 0 分散マルチ・トランザクション処理装置
- 9 0 「、9 「 「、9 2 「 データベース・オブジェクト・セット識別記号
変換部
- 9 0 2、9 「 2、9 2 2 データベース・オブジェクト・セット識別記号
ロケーション情報変換部
- 9 0 3 更新アクセス・カウンタ
- 9 0 5 データベース・オブジェクト・セット識別記号ロケーション情報
マップ・テーブル
- 9 0 6 分散マルチ・トランザクション処理実行部
- 9 0 7 データベース・オブジェクト管理装置特定装置
- 9 0 8 LAN、WANまたはマシン内のTCP/IPネットワーク、
またはATA、FC等のインターフェース・ケーブル

請求の範囲

[請求項1]

ネットワーク上に配置された複数のデータベース・ドメインの各々の前記データベース・ドメイン内にあって、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

自身のデータベース・ドメイン内のデータベース・オブジェクト管理装置が保持する管理下のデータベースの構造を定義するデータ・ディクショナリ情報と、

自身のデータベース・ドメイン内のデータベース・オブジェクト管理装置内のデータベース・オブジェクトが他の前記データベース・ド

メイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対して外部参照状態にある時の参照被参照状態にある双方のデータベース・オブジェクトに関するステータス情報と、

自身の前記データベース・ドメイン内の前記データベース・オブジェクト管理装置が管理する前記データベース・オブジェクトを他のデータベース・ドメインを管理する前記データベース・メタ情報管理リポジトリ装置を直接または間接的にアクセスする端末計算機またはサーバ計算機からの読み出し状態にあり共有ロック状態にある当該データベース・オブジェクトのステータス情報と、

自身の前記データベース・ドメイン内の前記データベース・オブジェクト管理装置が管理する前記データベース・オブジェクトを他のデータベース・ドメインを管理する前記データベース・メタ情報管理リポジトリ装置を直接または間接的にアクセスする端末計算機またはサーバ計算機からの書き込み状態にあり排他ロック状態にある当該データベース・オブジェクトのステータス情報と、

を分散型データベース・システムのデータベース・メタ情報として格納し、

自身のデータベース・ドメイン内で処理が進行し、前記データベース・オブジェクトに関するステータス情報に変化が及ぼす事象が発生した場合には速やかに当該データベース・オブジェクトに関するステータス情報を更新管理する、

データベース・メタ情報格納管理部と、

を有し、

前記分散データベースに参加中のデータベース・ドメイン内の「当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュ

をメモリ・イメージとして互いにノード間で共有することによって実現される、または、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部とが、前記「分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されることを特徴とするメタ情報共有型分散データベース・システム。

[請求項2]

ネットワーク上に配置された、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置と、

分散マルチ・トランザクション処理においてトランザクション処理一貫性を保持しながら実行するための分散マルチ・トランザクション処理装置と、

前記データベースの持つメタ情報を管理するための「または2以上のデータベース・メタ情報管理リポジトリ装置と、

で構成されるアクセス・カウンタ式分散マルチ・トランザクション制御システムであって、

データベース・メタ情報管理リポジトリ装置は、

「または2以上の更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別されることを特徴とする更新アクセス・カウンタと、

を有し、

分散マルチ・トランザクション処理装置は、

直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求されるデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトランザクション処理において、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、該起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドが存在すればそのプロセスないしスレッドによって要求される更新アクセスが発生する都度、該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを一定数ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致した時に、該前記起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのコミット処理を実行し、

また、実際の更新アクセスが失敗したという情報を一個でも受け取

った場合、余の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点となったプロセス自身および該前記起点となったプロセスないしスレッドが存在するならばそのプロセスないしスレッドによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのロールバック処理を実行する、

分散マルチ・トランザクション処理実行部と、

を有する

ことを特徴とする、アクセス・カウンタ式分散マルチ・トランザクション制御システム。

[請求項3]

ネットワーク上に配置された複数のデータベース・ドメインの各々の前記データベース・ドメイン内にあって、

「または2以上の請求項2に記載のアクセス・カウンタ式分散マルチ・トランザクション制御システムからで構成されるメタ情報共有分散型データベース・システムであって、

データベース・メタ情報管理リポジトリ装置は、さらに、外部の他のデータベース・ドメインと整合性をとるための機能を有し、

分散マルチ・トランザクション処理装置は、さらに、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求される、自身または他の前記データベース・ドメイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトランザクション処理において、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタが存在するものとし、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、

該起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドが存在すればそのプロセスないしスレッドによって要求される更新アクセスが発生する都度、該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを一定数ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致した時に、該前記起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのコミット処理を実行し、

また、実際の更新アクセスが失敗したという情報を一個でも受け取

った場合、余の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点となったプロセス自身および該前記起点となったプロセスないしスレッドが存在するならばそのプロセスないしスレッドによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないしスレッドによって実行されたすべての更新アクセスのロールバック処理を実行する、

ことを特徴とする分散マルチ・トランザクション処理実行部と、
を有し、

前記メタ情報共有分散型データベース・システムに参加中の前記データベース・ドメイン内の当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、または、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部とが、前記分散データベースに参加中のデータベース・ドメイン内の「各々のデータベース・メタ情報管理リポジトリ装置によって共有されること」を特徴とするメタ情報共有型分散データベース・システム。

[請求項4]

請求項「に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

分散マルチ・トランザクション処理においてトランザクション処理一貫性を保持しながら実行するための分散マルチ・トランザクション処理装置と、

を構成要素とするメタ情報共有分散型データベース・システムであって、

分散マルチ・トランザクション処理装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは当該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

直接的または間接的に、挿入または変更または削除の「または2以上の更新アクセスを端末計算機またはサーバ計算機から要求される、自身または他の前記データベース・ドメイン内のデータベース・オブジェクト管理装置が管理するデータベース・オブジェクトに対するトランザクション処理において、

複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタが存在するものとし、

前記更新アクセス・カウンタはあらかじめ特定の初期値を持つものとし、

該起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないし

スレッドが存在すればそのプロセスないスレッドによって要求される更新アクセスが発生する都度、該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを一定数ずつカウントアップし、

実際の更新アクセスが成功したという情報を順に受け取る毎に該前記更新アクセスを要求する起点となったプロセスを識別するプロセス識別子によって識別される該前記更新アクセス・カウンタ内のカウントを該前記一定数と同じ値ずつカウントダウンし、該前記カウンタの値が該前記初期値と一致した時に、該前記起点となったプロセス自身および該前記起点となったプロセスによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないスレッドによって実行されたすべての更新アクセスのコミット処理を実行し、また、

実際の更新アクセスが失敗したという情報を一個でも受け取った場合、余の更新アクセスを中止し、その結果を待たずに、すみやかに該前記起点となったプロセス自身および該前記起点となったプロセスないスレッドが存在するならばそのプロセスないスレッドによって直接または間接的に引き起こされた自身または他のデータベース・ドメイン内の端末計算機またはサーバ計算機上のプロセスないスレッドによって実行されたすべての更新アクセスのロールバック処理を実行する、

分散マルチ・トランザクション処理実行部と、

を有することを特徴とする分散マルチ・トランザクション処理実行部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ド

メインが複数個あるものとして、
複数のデータベース・ドメインから要求される「または2以上の前記更新アクセスの起点となるプロセスを識別するプロセス識別子によって識別される更新アクセス・カウンタ
を有すること
を特徴とするメタ情報共有型分散データベース・システム。

[請求項5]

ネットワーク上に配置された複数のデータベース・ドメインの各々の前記データベース・ドメイン内にあって、

データベースを管理する「または2以上のデータベース・オブジェクト管理装置と、

前記データベース・オブジェクト管理装置が管理するデータベースを格納するための、各々の前記データベース・オブジェクト管理装置毎に対応する「または2以上のデータベース・オブジェクト格納装置と、

単一の前記データベース・オブジェクト格納装置内に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、
ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替

えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブル

を有し、

前記メタ情報共有分散型データベース・システムに参加中の前記データベース・ドメイン内の当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、または、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の

変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部とが、前記分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されることを特徴とするメタ情報共有型分散データベース・システム。

[請求項6] 請求項「に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置内に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、

ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データ

ベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部

と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブル

を有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、または、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製す

ることによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部とが、前記「分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されること

を特徴とするメタ情報共有型分散データベース・システム。

[請求項7] 請求項3に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置内に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、

ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト

・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータベース・オブジェクト・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ドメインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記 分散データベースに参加中のデータベース・ドメイン内の」当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、または、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記デー

データベース・メタ情報格納管理部と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ装置内にあるデータベース・メタ情報格納管理部とが、前記「分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されることを特徴とするメタ情報共有型分散データベース・システム。

[請求項8]

請求項4に記載のメタ情報共有型分散データベース・システムにおいて、さらに、

単一の前記データベース・オブジェクト格納装置内に収容しきれないタプル数を持つリレーションにおいて、水平分割して複数のデータベース・ドメイン管理下の「または2以上の前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置内に収容しようとする際、対象データ・オブジェクトとデータ・オブジェクト管理装置を対応付けるためのデータベース・オブジェクト管理装置特定装置と、

複数のデータベース・ドメインにまたがるデータベース・オブジェクトの検索または加工または更新を行うためのデータベース・オブジェクト検索加工装置と、

前記データベースの持つメタ情報によって、外部の他のデータベース・ドメインと整合性をとるための「または2以上のデータベース・メタ情報管理リポジトリ装置と、

で構成されるメタ情報共有分散型データベース・システムであって

データベース・オブジェクト管理装置特定装置は、

ハッシング等のアルゴリズムを特定のパラメタで用いて、対象データベース・オブジェクトの一部ないし全部の情報を、対照データベース・オブジェクトの論理的な格納先となるデータベース・オブジェクト・セットを特定するためのデータベース・オブジェクト・セット識別記号に変換するデータベース・オブジェクト・セット識別記号変換部

と、

該前記データベース・オブジェクト・セット識別記号に対応する該前記データベース・オブジェクト・セットの物理的な格納先となる前記データベース・ドメインおよび前記データベース・オブジェクト管理装置を特定するデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報変換部と、

を有し、

データベース・オブジェクト検索加工装置は、

該前記アルゴリズムと同一のアルゴリズムを同一のパラメタで用いて、検索対象となる該前記データベース・オブジェクト・セットを特定しながら、「または2以上の該前記データベース・オブジェクト管理装置が管理する「または2以上の前記データベース・オブジェクト格納装置上の「または2以上のデータ・セットを横断的に検索するデータベース・オブジェクト検索部と、

前記データベース・オブジェクト検索部によって特定され抽出されたデータを、該前記アルゴリズムと同一のアルゴリズムを同一のパラメタを用いて、加工対象となる該前記データベース・オブジェクト・セットを特定しながら、変更または削除、若しくは「または2以上の該前記データベース・オブジェクト・セットにまたがるデータ群の結合すなわちジョインまたは射影すなわちプロジェクションまたは並べ替えすなわちソートまたは行結合すなわちユニオン等のデータベース・オブジェクト加工を横断的に行うデータベース・オブジェクト加工部と、

を有し、

データベース・メタ情報管理リポジトリ装置は、

ここではネットワーク上には互いに通信が可能なデータベース・ド

メインが複数個あるものとして、また、

ここではデータベース・オブジェクトは該データベース・オブジェクトを管理する前記データベース・オブジェクト管理装置を識別するための情報であるデータベース・オブジェクト管理装置識別子と当該データベース・オブジェクト管理装置を管理する前記データベース・ドメインを識別するためのデータベース・ドメイン識別子とに関連付けられたデータベース・オブジェクト識別子で識別されるものとして、

前記データベース・オブジェクト管理装置特定装置の前記データベース・オブジェクト・セット識別記号ロケーション情報変換部において保持され適時更新されるデータベース・オブジェクト・セット・ロケーション情報と該前記データベース・オブジェクト・セット識別記号とを対応付けするデータベース・オブジェクト・セット識別記号ロケーション情報マップ・テーブルを有し、

前記メタ情報共有分散型データベース・システムに参加中の前記データベース・ドメイン内の当該データベース・メタ情報管理リポジトリ装置を実装するためのサーバ計算機である各ノードの「主記憶装置の一部」をメモリ・キャッシュとし、自身のノードおよび他のノード上の前記メモリ・キャッシュをメモリ・イメージとして互いにノード間で共有することによって実現される、または、前記メモリ・キャッシュを互いにノード間で複製し、保持し、データベース・メタ情報の変化に伴い、該変化にかかわる該ノード上の該変化にかかわる該メモリ・キャッシュ上のメモリ・イメージを他のノード上の対応するメモリ・キャッシュ上に複製することによって同期が行われ実現される、単一の仮想主記憶装置を構成し、自身のデータベース・メタ情報管理リポジトリ装置の前記データベース・メタ情報格納管理部と、他のデータベース・ドメイン内の前記データベース・メタ情報管理リポジトリ

リ装置内にあるデータベース・メタ情報格納管理部とが、前記分散データベースに参加中のデータベース・ドメイン内の」各々のデータベース・メタ情報管理リポジトリ装置によって共有されることを特徴とするメタ情報共有型分散データベース・システム。

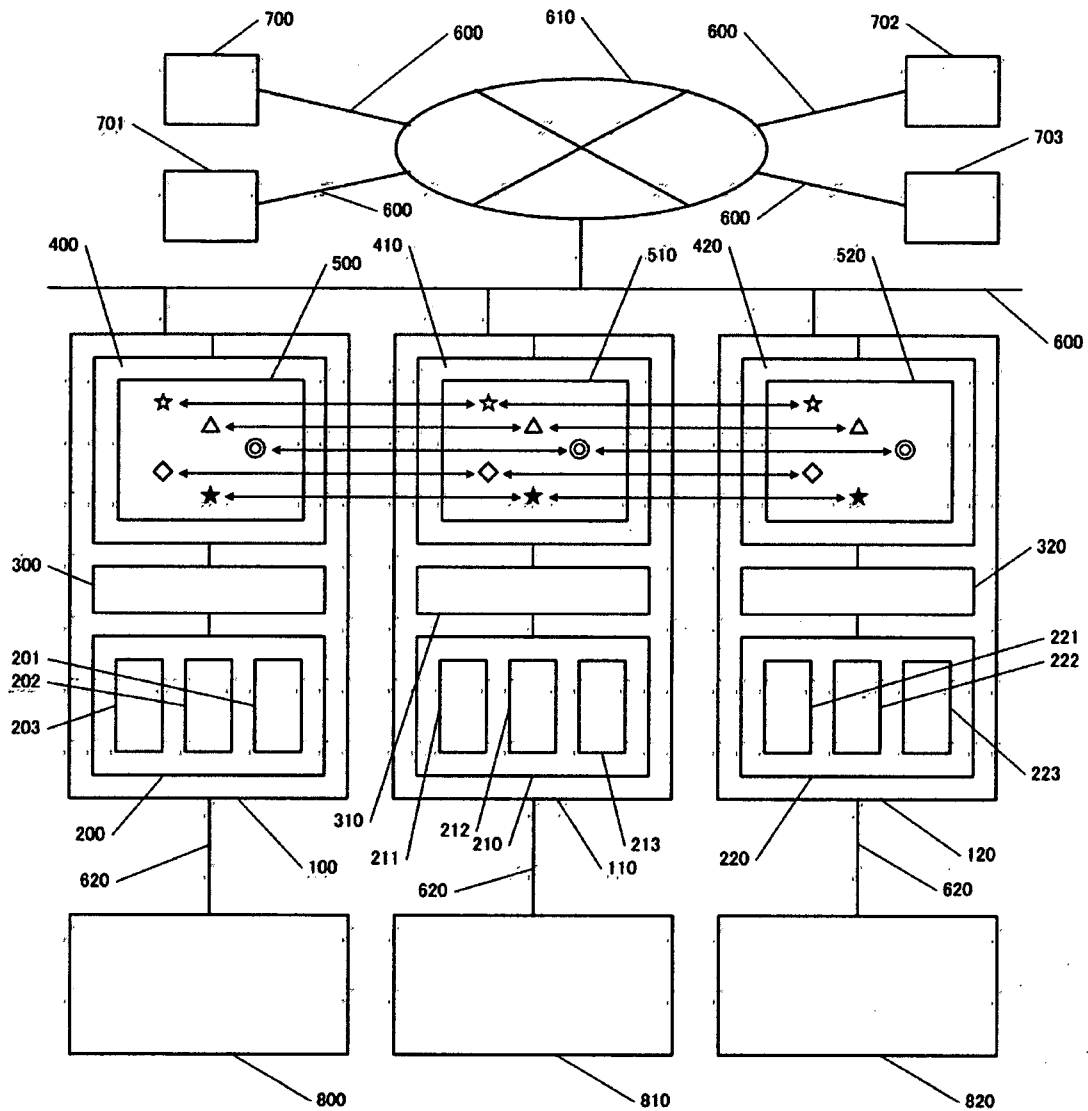
[請求項9] 前記データベース・オブジェクト管理装置特定装置は、前記データベース・オブジェクト・セット識別記号ロケーション情報変換部のロケーション情報を変更する機能を有し、前記データベース・オブジェクト格納装置の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする請求項5に記載のメタ情報共有型分散データベース・システム。

[請求項10] 前記データベース・オブジェクト管理装置特定装置は、前記データベース・オブジェクト・セット識別記号ロケーション情報変換部のロケーション情報を変更する機能を有し、前記データベース・オブジェクト格納装置の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする請求項6に記載のメタ情報共有型分散データベース・システム。

[請求項11] 前記データベース・オブジェクト管理装置特定装置は、前記データベース・オブジェクト・セット識別記号ロケーション情報変換部のロケーション情報を変更する機能を有し、前記データベース・オブジェクト格納装置の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする請求項7に記載のメタ情報共有型分散データベース・システム。

[請求項12] 前記データベース・オブジェクト管理装置特定装置は、前記データベース・オブジェクト・セット識別記号ロケーション情報変換部のロケーション情報を変更する機能を有し、前記データベース・オブジェクト格納装置の記憶装置容量と格納されるデータベース・オブジェクト・セットの大きさの関係を最適化して再配置することを特徴とする請求項8に記載のメタ情報共有型分散データベース・システム。

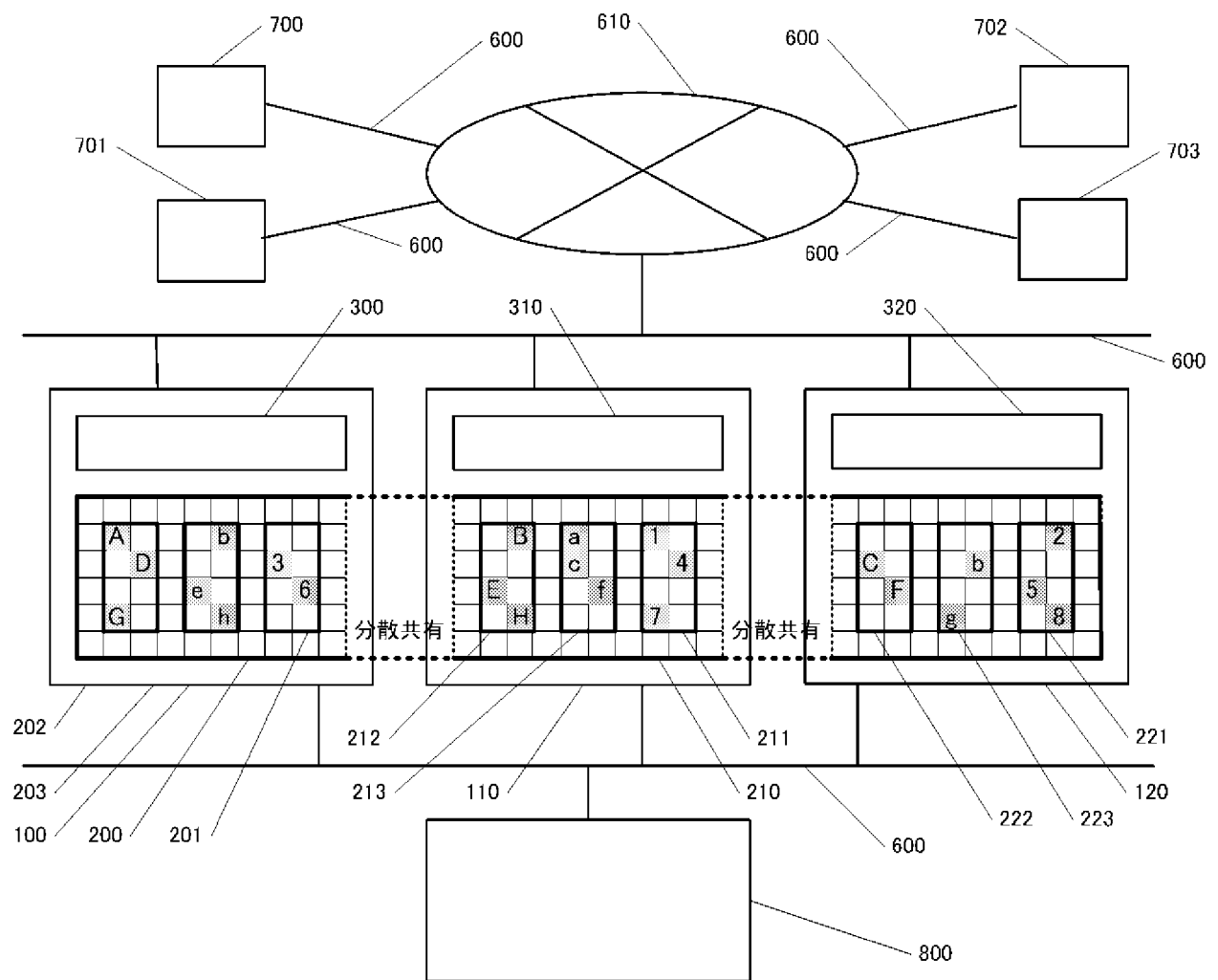
[図 1]

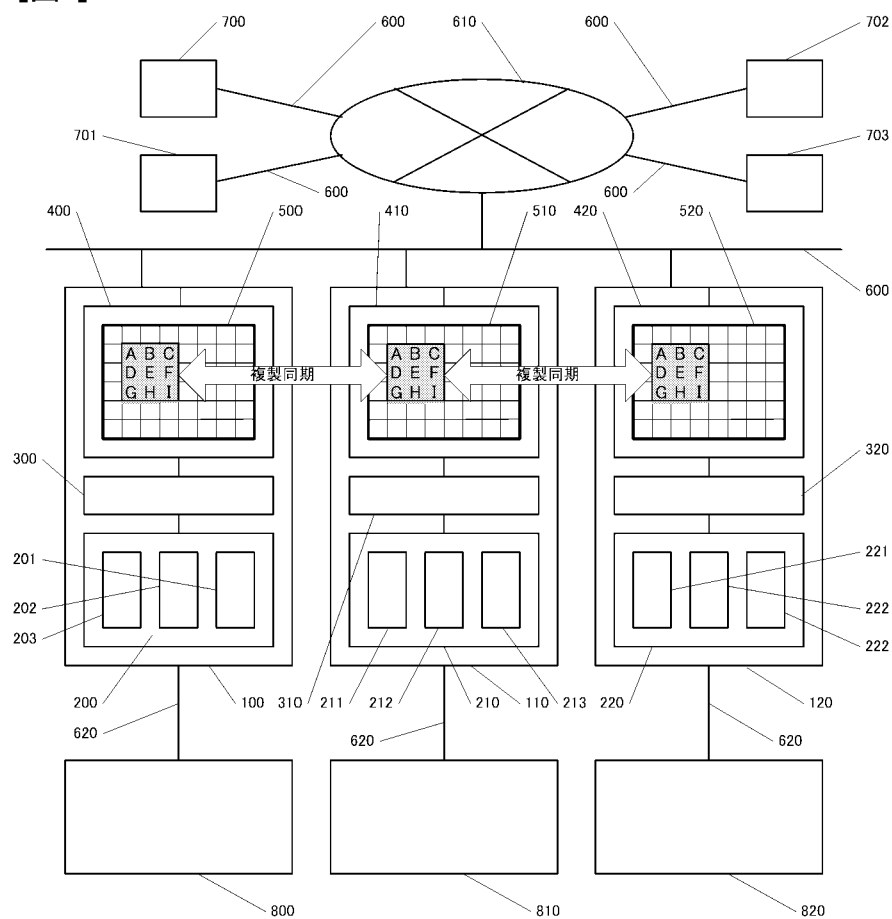


凡例

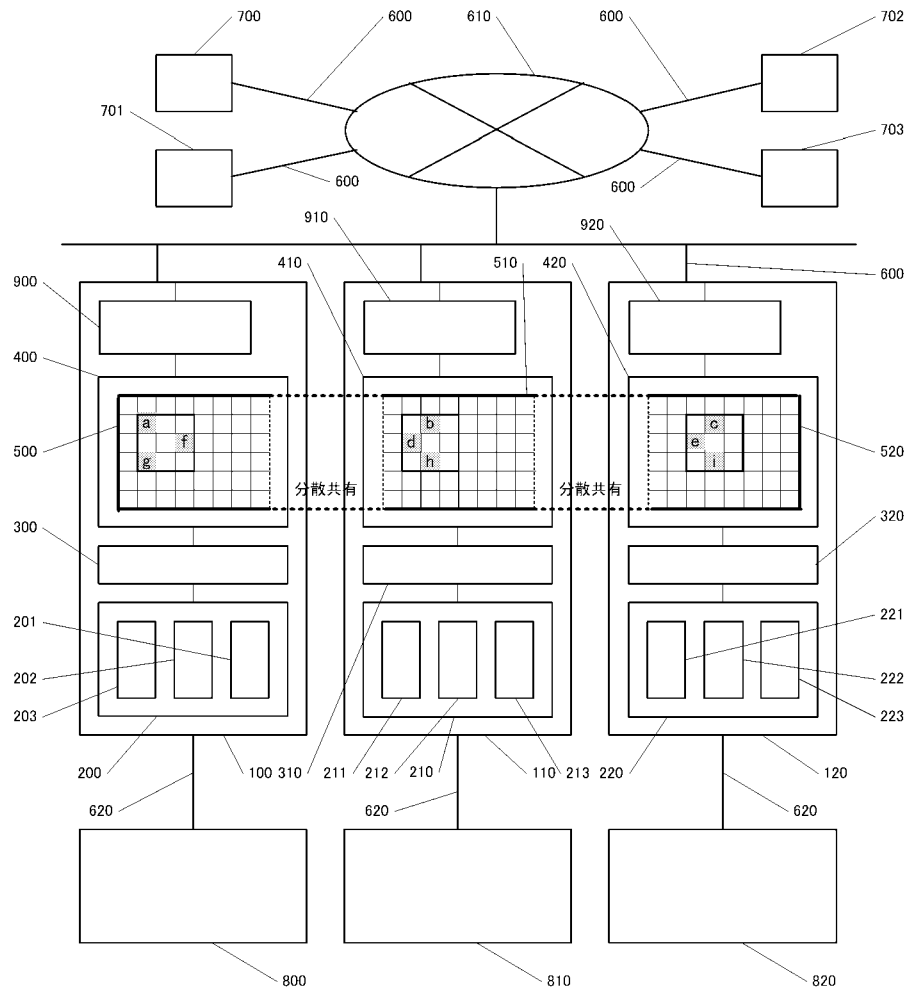
- ☆ ↔ ☆ 共有ロックステータス各個同期
- △ ↔ △ 排他ロックステータス各個同期
- ◎ ↔ ◎ 参照整合性ステータス各個同期
- ◇ ↔ ◇ 更新カウンタ各個同期
- ★ ↔ ★ マップ・テーブル各個同期

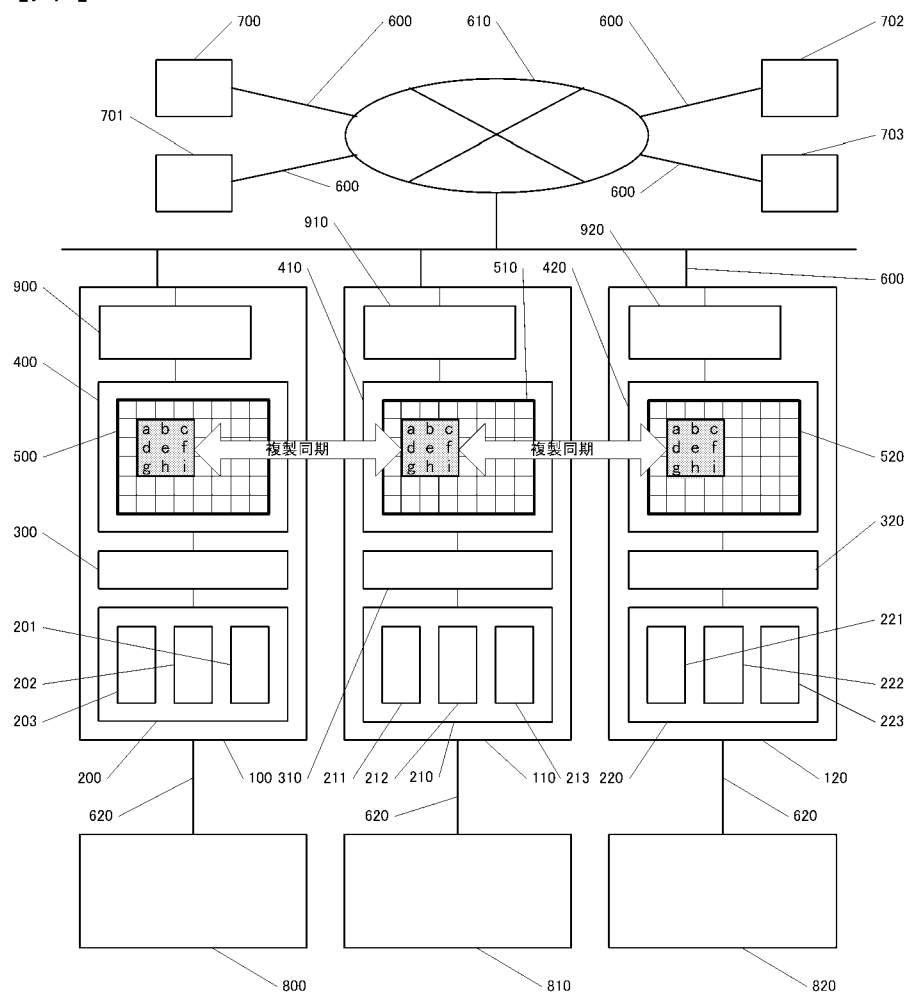
[図2]



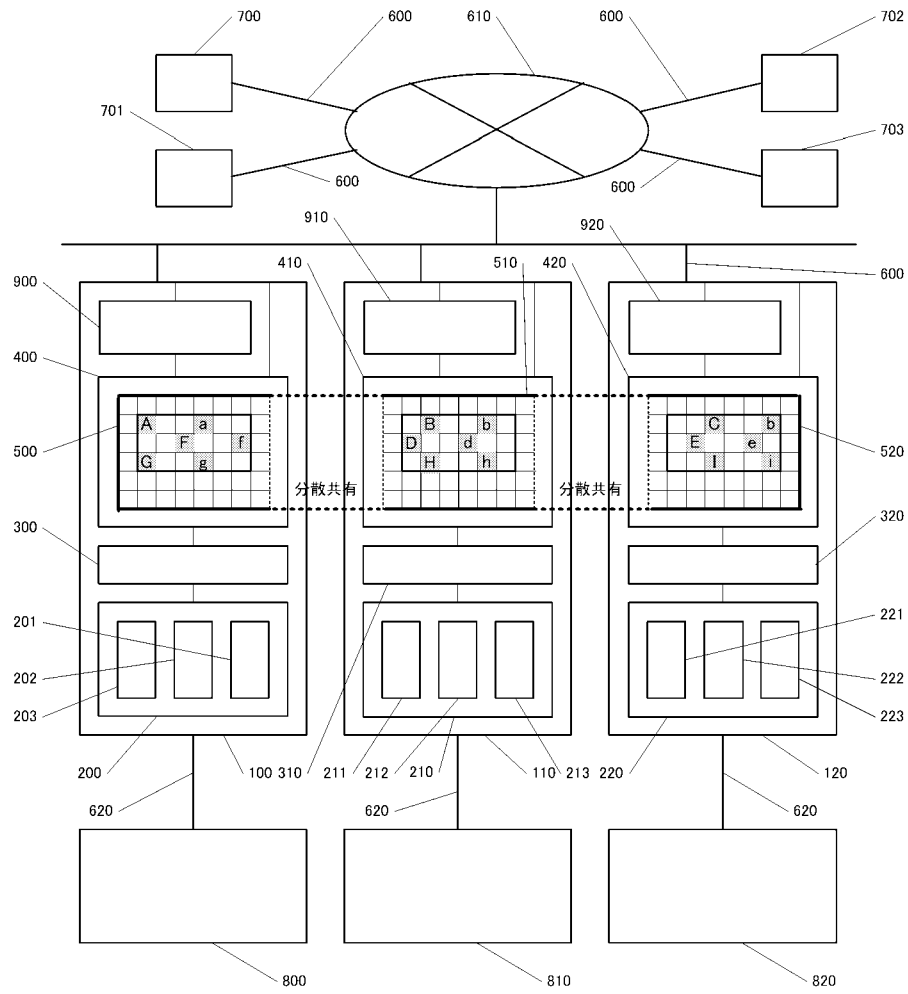


[図5]

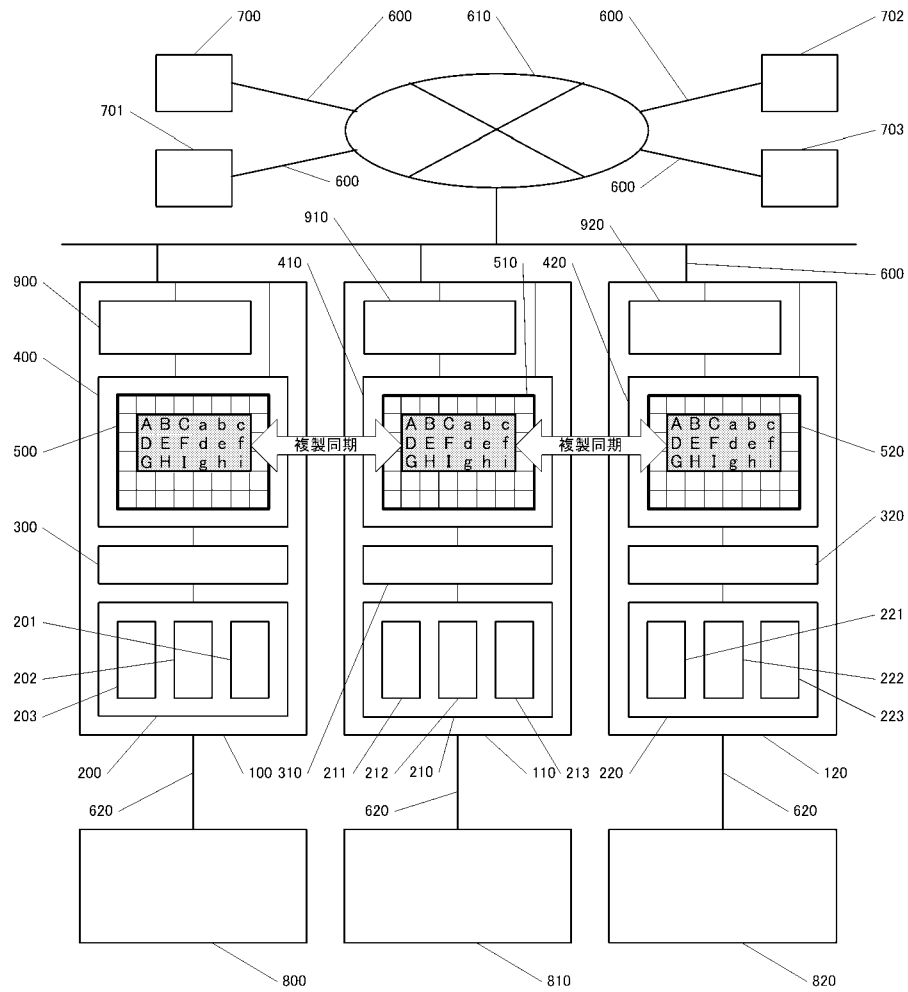




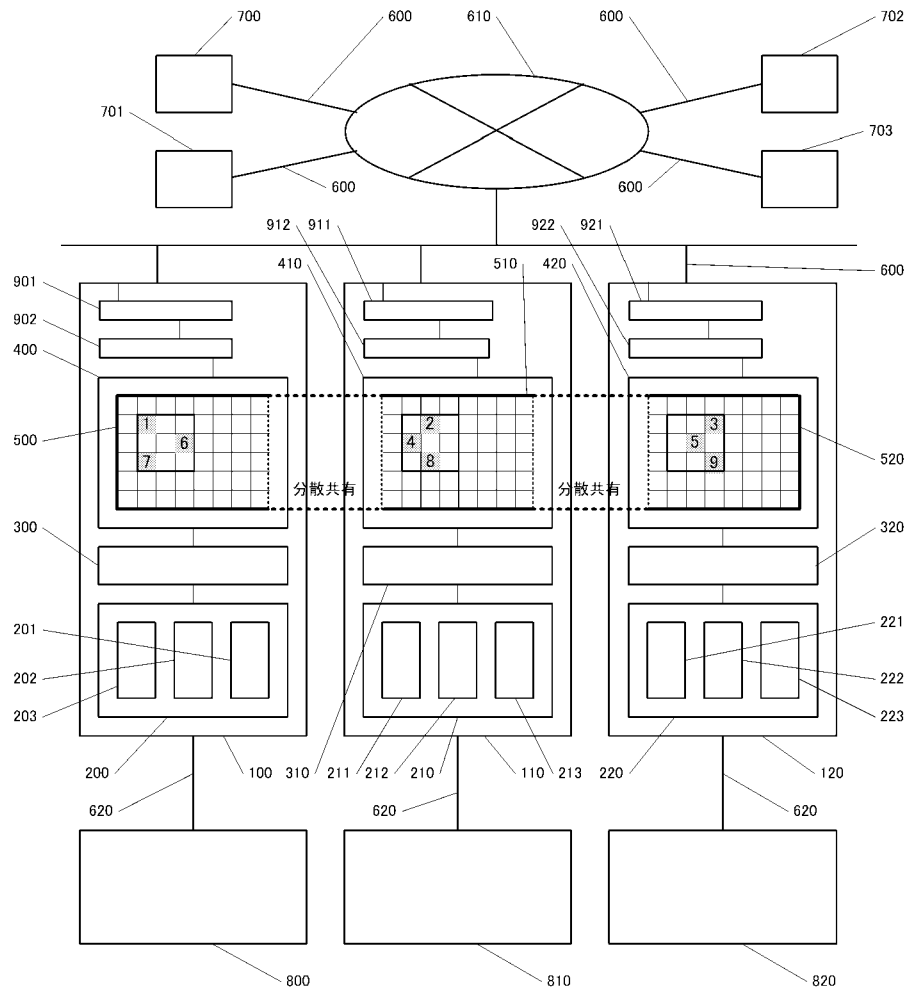
[図7]



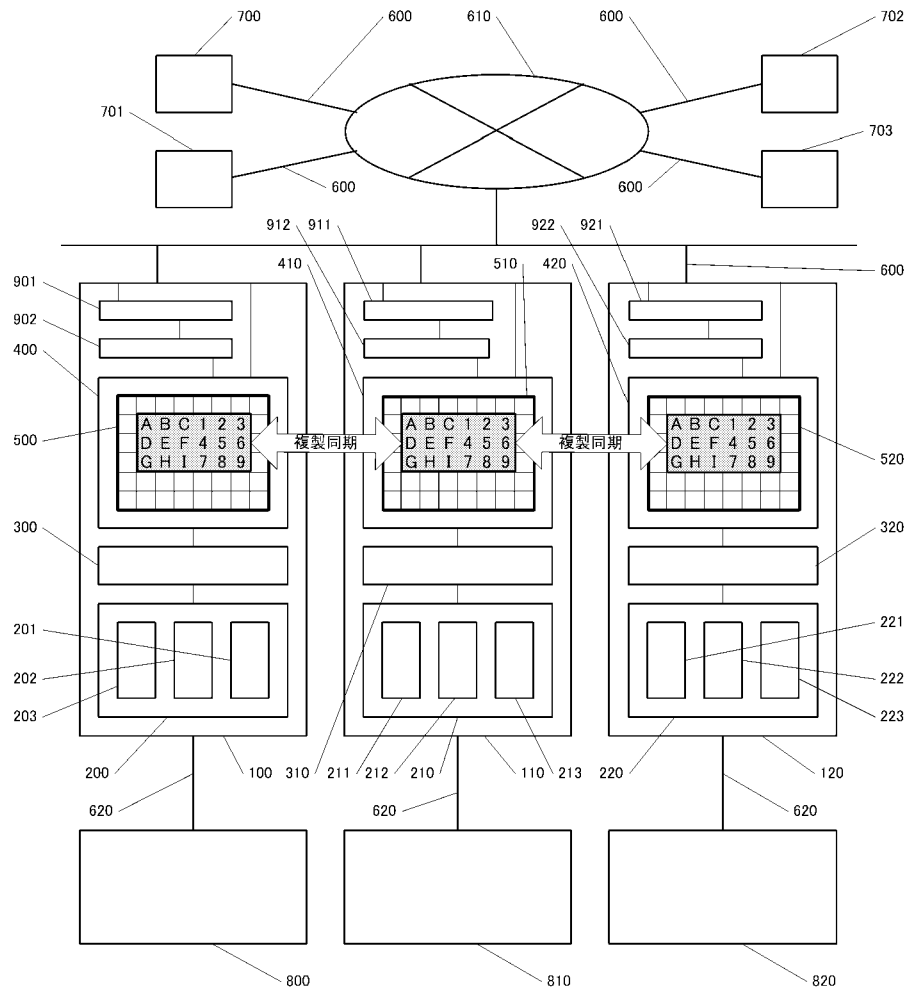
[図8]



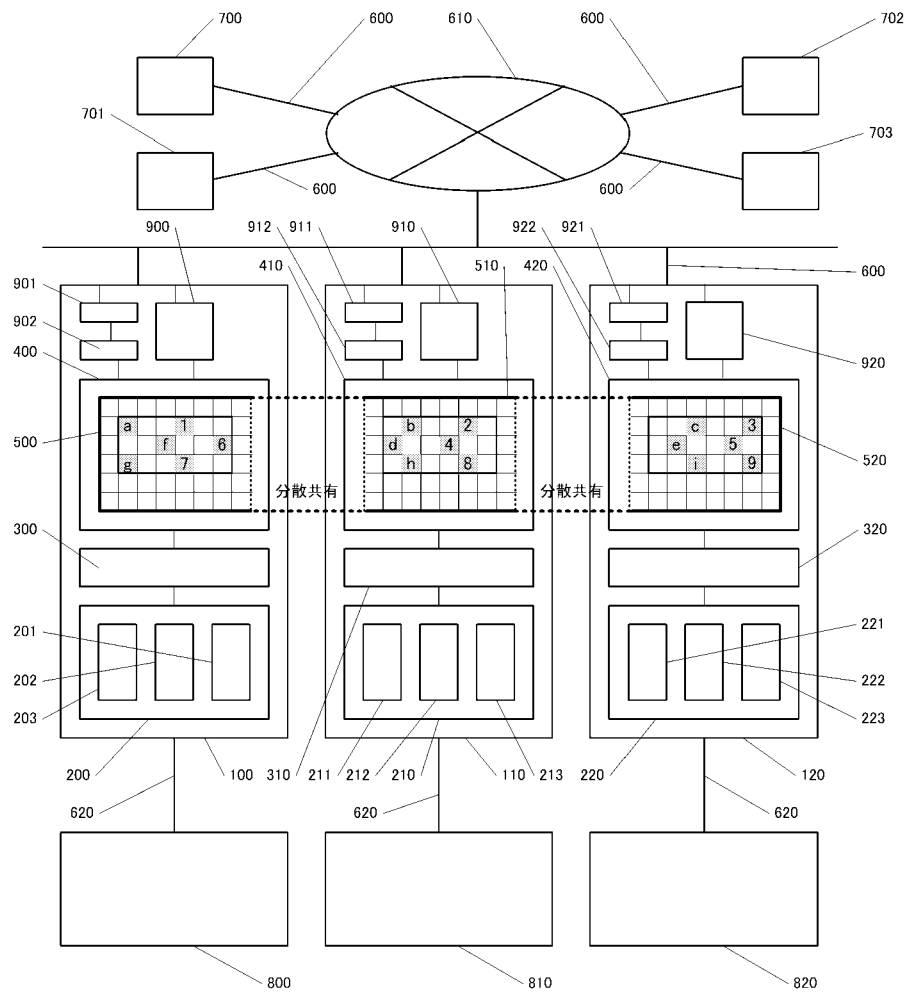
[図9]

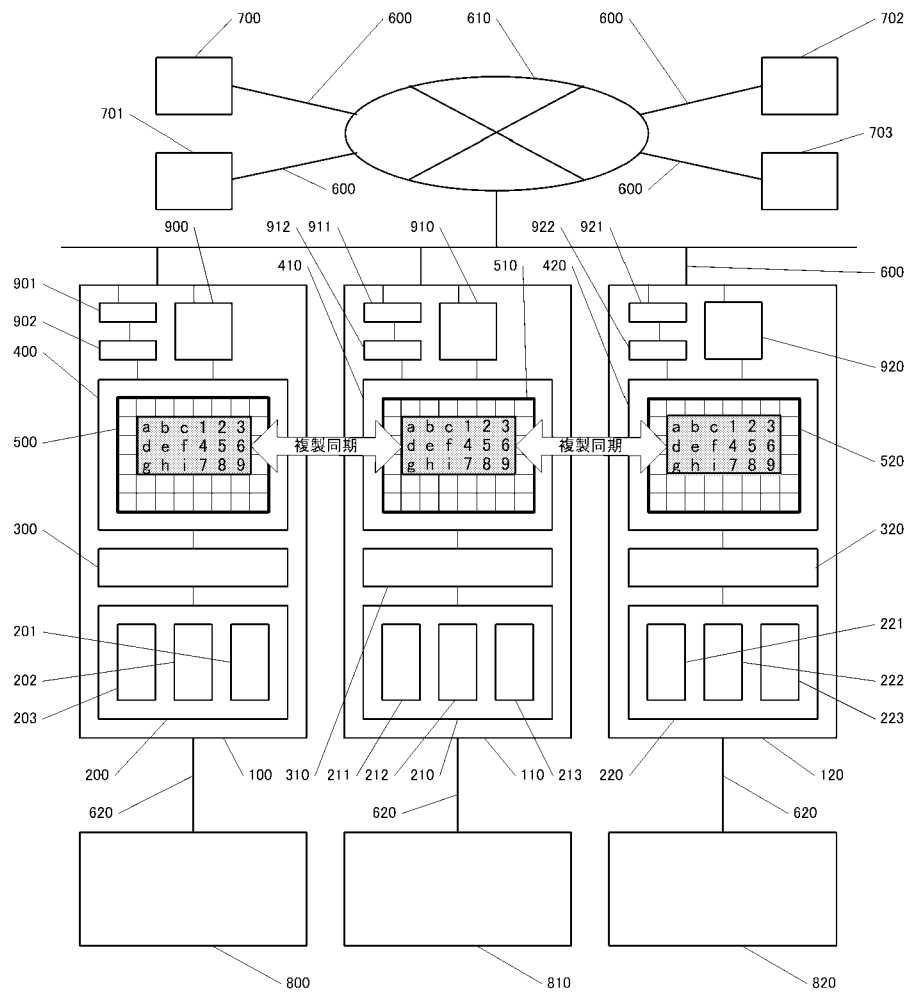


[図12]

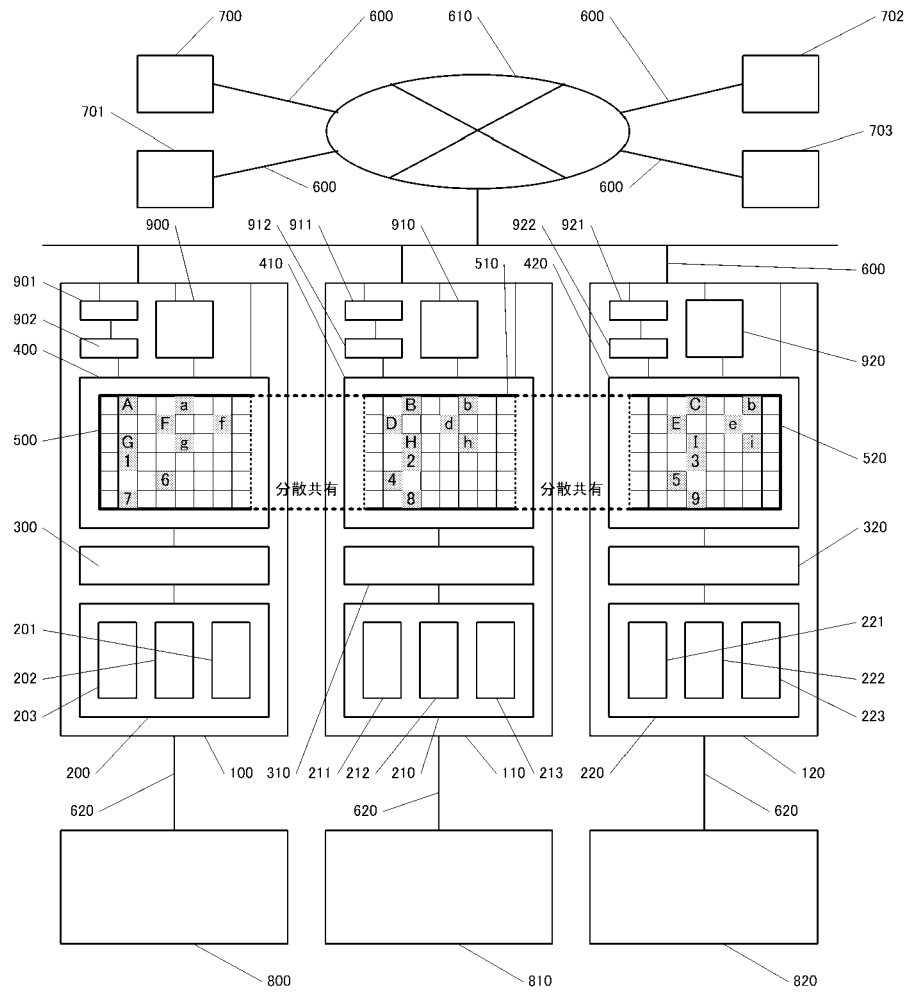


[図13]

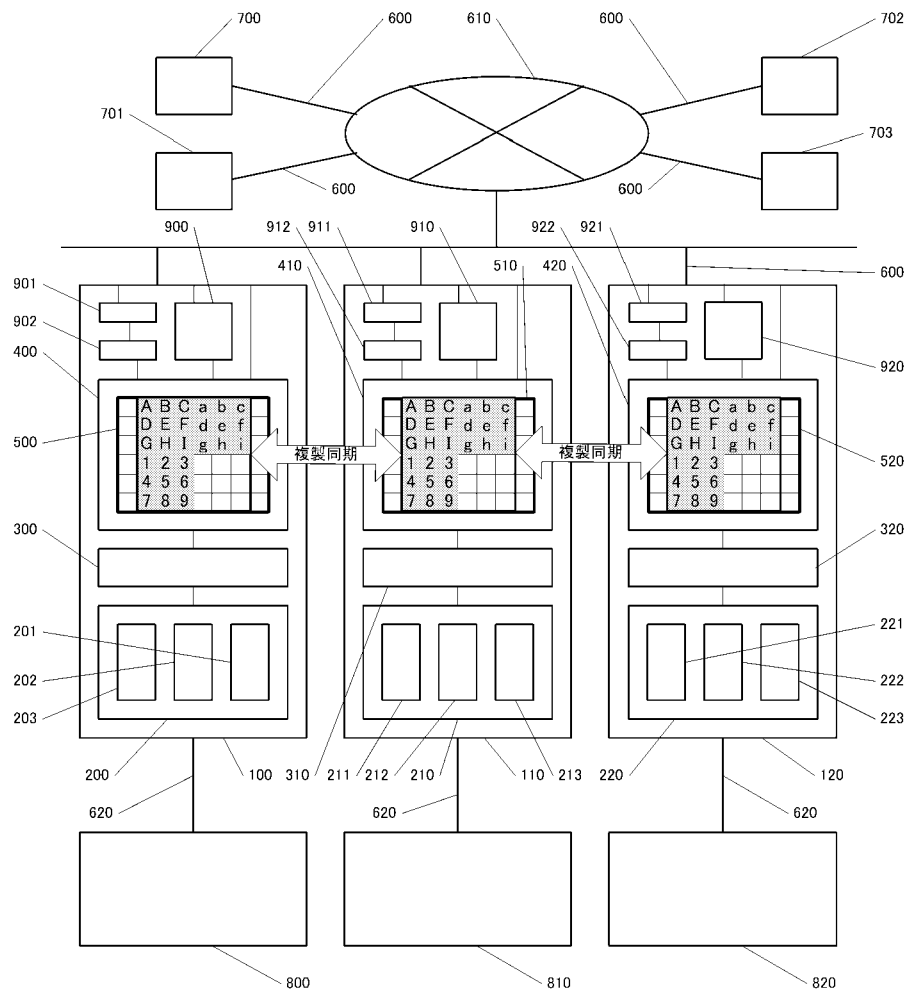




[図15]



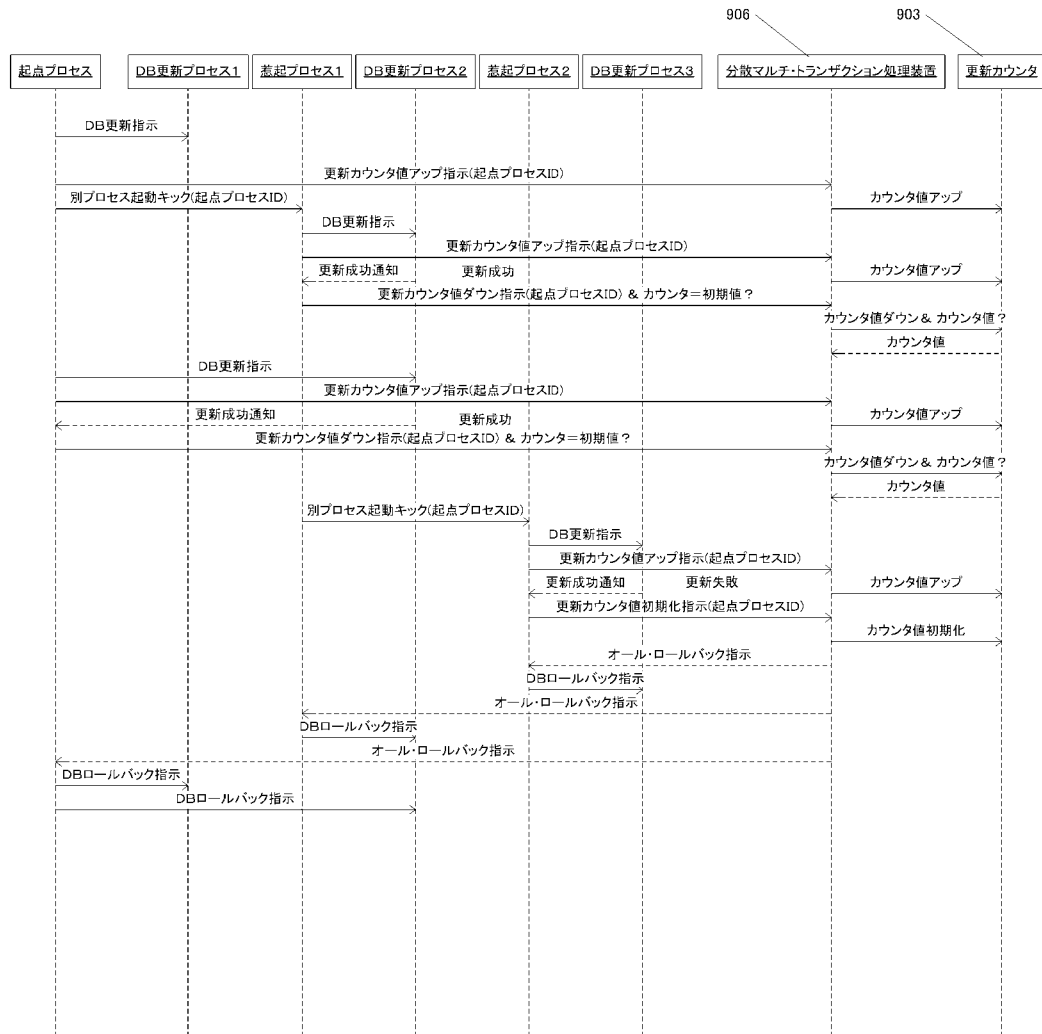
[図16]



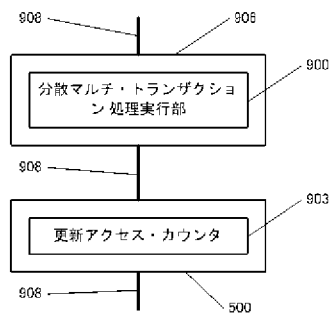
[図17]



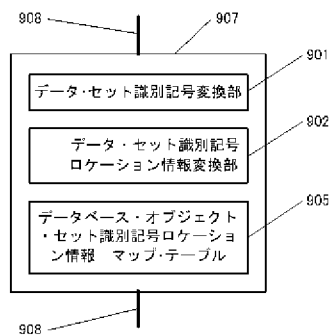
[図18]



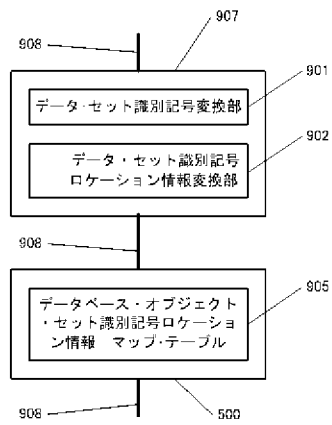
[図19]



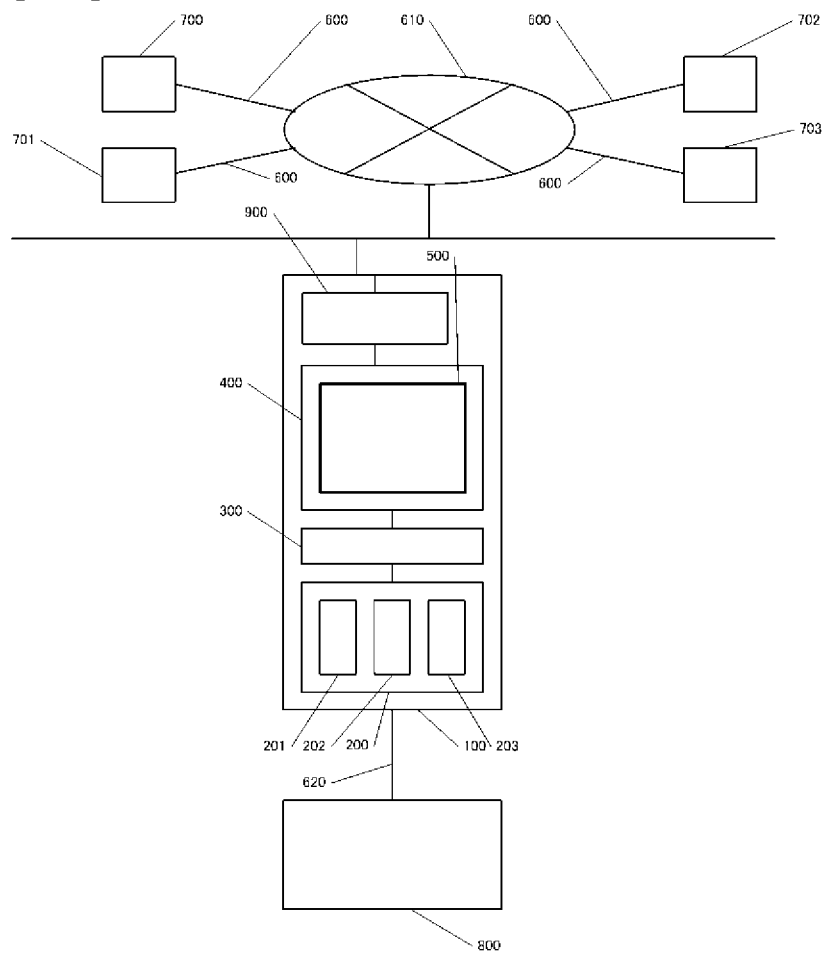
[図20]



[図21]



[図22]



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2 010/057 886

A. CLASSIFICATION OF SUBJECT MATTER

G06F12/00 (2006.01) ±

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F12/00

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2010
Kokai Jitsuyo Shinan Koho	1971-2010	Toroku Jitsuyo Shinan Koho	1994-2010

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

IEEE Xplore, JSTPlus (JDreaml l)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2007-501456 A (Oracle International Corp.), 25 January 2007 (25.01.2007), paragraphs [0002] to [0016] & WO 2005/013155 A1	1-12
A	JP 2002-342138 A (The Hachijuni Bank Ltd.), 29 November 2002 (29.11.2002), paragraphs [0002] to [0018] (Family: none)	1-12
A	JP 2007-219598 A (Nippon Telegraph And Telephone Corp.), 30 August 2007 (30.08.2007), claim 1 (Family : none)	1-12



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

07 June , 2010 (07.06.10)

Date of mailing of the international search report

22 June , 2010 (22.06.10)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2 010/057 886

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2004-227169 A (Shuhei NISHIYAMA) , 12 August 2004 (12.08.2004) , claim 1 & WO 2004/066152 A1	1-12
A	JP 2006-092503 A (Shuhei NISHIYAMA) , 06 April 2006 (06.04.2006) , claim 1 (Family: none)	1-12

A. 発明の属する分野の分類 (国際特許分類 (IPC))

Int Cl G06F12/00 (2006. 01) i

B. 調査を行った分野

調査を行った最小限資料 (国際特許分類 (IPC))

IntCl G06F12/00

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1922-1996 年
日本国公開実用新案公報	1971-2010 年
日本国実用新案登録公報	1996-2010 年
日本国登録実用新案公報	1994-2010 年

国際調査で利用した電子データベース (データベースの名称、調査に使用した用語)

IEEE Xplore, JSTPlus (JDreamII)

C. 関連すると認められる文献

引用文献の カテゴリーホ	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	JP 2007-501456 A (オラクル・インターナショナル・コーポレーション) 2007. 01. 25, 段落 0002-0016 & wo 2005/013155 A1	1-12
A	JP 2002-342138 A (株式会社八十二銀行) 2002. 11. 29, 段落 0002-0018 {Y アミリーなし}	1-12
A	JP 2007-219598 A (日本電信電話株式会社) 2007. 08. 30, 請求項 1 (7 アミリーなし)	1-12

洋 C 欄の続きにも文献が列挙されている。

ヴ パテントファミリーに関する別紙を参照。

ホ 引用文献のカテゴリー	の日の役に公表された文献
「A」特に関連のある文献ではなく、一般的技術水準を示すもの	IT」国際出願日又は優先日後に公表された文献であつて出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの
IE」国際出願日前の出願または特許であるか、国際出願日以後に公表されたもの	IX」特に関連のある文献であつて、当該文献のみで発明の新規性又は進歩性がないと考えられるもの
「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献 (理由を付す)	IY」特に関連のある文献であつて、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの
「O」口頭による開示、使用、展示等に言及する文献	r&j 同一パテントファミリー文献
「p」国際出願日前で、かつ優先権の主張の基礎となる出願	

国際調査を完了した日 07. 06. 2010	国際調査報告の発送日 22. 06. 2010
国際調査機関の名称及びあて先 日本国特許庁 (ISA/JP) 郵便番号 100-8915 東京都千代田区霞が関三丁目4番3号	特許庁審査官 (権限のある職員) 田川 泰宏 電話番号 03-3581-1101 内線 3599
	5M 4236

C (続き) . 関連すると認められる文献		
引用文献の テコリーホ	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求大項の番号
A	JP 2004-227169 A (西山修平) 2004. 08. 12, 請求項 1 & WO 2004/066152 A1	1-12
A	JP 2006-092503 A (西山修平) 2006. 04. 06, 請求項 1 (7 アミ Vーなし)	1-12